

## RESEARCH OUTPUTS / RÉSULTATS DE RECHERCHE

### Un service d'archivage pour les sciences sociales en Belgique ?

Laurent, Véronique; Lefever, Nathalie; Schoups, Inge; Lobet-Maris, Claire; Poulet, Yves

*Publication date:*  
2007

*Document Version*  
Première version, également connu sous le nom de pré-print

[Link to publication](#)

*Citation for published version (HARVARD):*

Laurent, V, Lefever, N, Schoups, I, Lobet-Maris, C & Poulet, Y 2007, *Un service d'archivage pour les sciences sociales en Belgique ? Etude de Faisabilité. Rapport pour le SPF Politique Scientifique Fédérale, décembre 2007*. CRID, Namur.

#### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

#### Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



## **Studie SODA**

### **Haalbaarheid van een data-archief voor de sociale wetenschappen**

## **Etude Prospect SODA**

### **Faisabilité d'un Service d'archivage de données pour les sciences sociales**

Studie uitgevoerd in 2007  
Rapport afgewerkt voorjaar 2008

In opdracht van 



## INHOUD

INHOUD .....	3
INLEIDING.....	5
 1ère PARTIE : ANALYSE DE L'EXISTANT.....	11
Introduction : Méthodologie .....	11
Chapitre 1 : Archivage en Belgique.....	13
Chapitre 2 : Analyse des expériences étrangères .....	17
Chapitre 3. Le contexte international .....	34
 2ème PARTIE : BESOINS ET ATTENTES DES ACTEURS AUTOUR DE L'ARCHIVAGE .....	39
Introduction : Méthodologie .....	39
Chapitre 1. Groupe de discussion avec les acteurs institutionnels.....	39
Chapitre 2. Groupe de discussion avec les acteurs scientifiques et académiques .....	41
 3ème PARTIE : MISE EN PLACE DU SERVICE D'ARCHIVAGE .....	43
Introduction.....	43
Chapitre 1. Un portail ou un service d'archivage ? .....	43
Chapitre 2. Types de données dans le service d'archivage.....	46
Chapitre 3. Dépôt de données dans le service d'archivage.....	49
Chapitre 4. Accès et dynamique de gestion des données.....	53
Hoofdstuk 5. Technische voorwaarden voor archivering .....	57
Chapitre 6. Modalités institutionnelles de création du service .....	73
Chapitre 7. Modalités juridiques du dépôt et de la réutilisation des données issues de la recherche.....	78
Technisch besluit.....	87
Conclusions générales.....	89
 BIJLAGES / ANNEXES.....	93
BIJLAGE A. Feasibility study of a Belgian computerised Archive Service for the social sciences (SODA). English description of the project .....	93
BIJLAGE B. Bibliografie .....	95
 ANNEXE 1. Liste de centres et de portails d'archives de données en sciences sociales .....	100
ANNEXE 2. Grille d'analyse sociologique des portails d'archives en sciences sociales.....	102
ANNEXE 3. Liste de questions juridiques pour l'établissement d'une grille d'analyse des portails en sciences sociales.....	104

ANNEXE 4. Exemple de questionnaire utilisé pour les visites des centres d'archives.....	105
ANNEXE 5. Programme de discussion du groupe institutionnel (27 juin 2007) .....	113
ANNEXE 6. Programme de discussion des représentants des universités et de la recherche (20 juillet 2007) .....	118
ANNEXE 7. Exemples de contrats .....	124
Chapitre 1. Exemples de licences de dépôt de données au centre d'archivage .....	124
Chapitre 2. Exemples de contrats de réutilisation des données .....	132
ANNEXE 8. International workshop on the creation of a Belgian Social Data Archive – Programme (16 october 2007) .....	139
ANNEXE 9. Règlement général en matière de propriété, de protection et de valorisation des résultats des recherches réalisées au sein des FUNDP .....	140

## INLEIDING

### **Een haalbaarheidsstudie voor de organisatie of inrichting van een data-archief voor de sociale wetenschappen.<sup>1</sup>**

**Werktitel: SODA – sociaal data-archief**

#### Algemene doelstellingen van het onderzoek

Sociale en andere wetenschappers verzamelen empirisch materiaal waarop ze hun analyses baseren. Doorgaans blijven deze gegevens na het afsluiten van het onderzoek ongebruikt liggen en dreigen voor altijd verloren te gaan. Gelet op de investeringen die vereist zijn om deze data te produceren en de vaak unieke wetenschappelijke waarde die deze bronnen bevatten, lijkt het aangewezen om deze reeksen toe te vertrouwen aan een instelling (data-archief) die de bronnen archiveert en ter beschikking stelt.

Onder data-archief wordt in deze context verstaan het toegankelijk maken en beschikbaar stellen van digitale databestanden in het domein van de maatschappijwetenschappen.

De databestanden waarvan sprake hebben betrekking op onderzoeksgegevens die wetenschappers en vorsers in de loop van hun onderzoek hebben opgesteld en die bij het archief worden gedeponereerd.

Tot de mogelijke taken van het data-archief behoren:

- het traceren van de beschreven databestanden,
- het conserveren van de gedeponeerde bestanden,
- het opstellen van metadata van de beschikbare collecties,
- het hergebruik van de databestanden verzekeren,
- het uniformiseren volgens internationale standaarden,
- het bemiddelen met buitenlandse data-archieven en bestanden

Onderstaande onderwerpen werden als onderzoeksvragen geformuleerd:

---

<sup>1</sup> Zie de bijlage (Annexe/bijlage A) voor een Engelse samenvatting

1. een beknopte beschrijving van de behoeften onder ambtenaren en onderzoekers (via een beperkt aantal interviews)
2. het in kaart brengen wat in dit domein reeds werd verwezenlijkt in België en in het buitenland
3. een bevraging van potentieel geïnteresseerde instellingen bij dit initiatief en nagaan hoe ze concreet betrokken kunnen worden bij de totstandkoming van het data-archief
4. een voorstel hoe het data-archief op structurele wijze kan gevoed worden, eventueel via de lopende onderzoeksprogramma's.

De werkopdracht werd uitgesplitst in 10 taken

#### Taak 1.

*Une typologie des bases et des données utilisées par les sciences humaines et sociales /*

Een typologie van de databestanden en data die worden gebruikt in de humane en de sociale wetenschappen.

#### Taak 2

*Une analyse des exigences des chercheurs et de leurs institutions vis à vis de ce projet menée via des interviews auprès d' équipes de recherches universitaires (francophones et néerlandophones) représentatives de la variété des situations (bases et données) identifiées au point 1 /* Een analyse van de behoeften bij onderzoekers en onderzoeksinstellingen met betrekking tot dit project door middel van interviews bij universitaire onderzoeksploegen (Frans- en Nederlandstaligen) die vertrouwd zijn met uiteenlopende data en databanken, zoals geïdentificeerd in taak 1.

#### Taak 3

*Une analyse de projets analogues menés en Belgique et à l'étranger basée d'une part sur une analyse "en ligne" d'un ensemble de plateformes ou archives existantes en sciences humaines et sociales (structure, accès, protection des données, propriété intellectuelle, etc.) devant servir à sélectionner des "best practices" et d'autre part sur des interviews en profondeur menées auprès des gestionnaires de plateformes "best practices" /* Een analyse van soortgelijke projecten in België en het buitenland door middel van enerzijds een webanalyse van bestaande data-archieven in de humane en sociale wetenschappen (structuur, toegang, bescherming van gegevens, intellectuele eigendom, etc.) met het oog op de selectie van "best practices" en anderzijds door middel van diepte-interviews bij beheerders van de "best practices".

#### Taak 4

*Interviews auprès d'institutions potentiellement intéressées par cette initiative et examen de leur implication concrète dans la construction d'une data archive (institutions fédérales, communautaires finançant la recherche scientifique, gestionnaires de collections de données) /* Bevragen van potentieel geïnteresseerde instellingen bij dit initiatief en nagaan hoe ze concreet betrokken kunnen worden bij de totstandkoming van het data-archief (federale instellingen en Gemeenschappen : instellingen die wetenschappelijk onderzoek financieren, beheerders van datacollecties).

#### Taak 5

Description de l'organisation et la structure de la data archive sur base de la réponse aux questions suivantes :

- *Quelles dispositions sont à prendre pour mettre sur pied un Service d'archivage? Quelles sont les contraintes institutionnelles et légales qui entravent éventuellement l'existence d'un Service d'archivage (protection des données, propriété intellectuelle)? Sous quelle forme ou statut cette initiative doit-elle se présenter?*

- *Quelle est la base réglementaire nécessaire pour assurer le dépôt et la consultation de collections de données dans le Service d'archivage?*
- *Quelle financement doit être considéré à cet égard ? /*

Een beschrijving van de organisatie en de structuur van het data-archief aan de hand van antwoorden op volgende vragen :

- welke maatregelen zijn er nodig om het data-archief in het leven te kunnen roepen? Wat zijn de institutionele en legale hindernissen die de realisatie eventueel in de weg staan (bescherming van auteursrecht en van de persoonlijke levenssfeer)? onder welke vorm/statuut is het initiatief concreet realiseerbaar?
- wat is de reglementaire basis voor het deponeren en het raadplegen van collecties in het data-archief ?
- welke financiering dient hiervoor gezocht te worden?

#### Taak 6

*Description des stratégies d'archivage et des métadonnées permettant la consultation par des tiers ; description du profil des gestionnaires de ce Service d'archivage /* Beschrijving van de archiveringsstrategieën en van de metadata nodig voor raadpleging door derden ; beschrijving van het profiel van de beheerders van het data-archief.

#### Taak 7

*Description des conditions nécessaires pour la consultation et l'usage de bases de données par des tiers /* Beschrijving van de voorwaarden van raadpleging en gebruik van de databestanden door derden.

#### Taak 8

*Organisation d'un séminaire pour présenter l'état de la connaissance scientifique au niveau international dans le domaine /* Organisatie van een seminarie waar de stand van de wetenschappelijke kennis op internationaal vlak wordt voorgesteld.

#### Taak 9

*Proposition de composition du comité d'accompagnement ; préparation des réunions; rapportages /* Voorstel van samenstelling van het comité; voorbereiding van vergaderingen; verslagen

#### Taak 10

*Les résultats de cette étude seront consignés dans un rapport final où les questions citées plus haut seront traitées et qui délivrera une information utile pour la mise sur pied d'un Service d'archivage*

*Le rapport contient e.a.:*

- *une description de la manière dont la constitution de la data archive doit être réglée sur le plan juridique ;quelles garanties doivent être rencontrées*
- *quelles sont les exigences techniques et les implications financières*
- *une description des facteurs critiques de succès ; les stakeholders et les partenaires*
- *formulaires, modèles de contrat, modèle de métadonnées /*

De studie zal resulteren in een eindrapport dat een antwoord zal bieden op hoger beschreven vragen en dat bruikbare en concrete informatie zal opleveren voor de oprichting van een sociaal data-archief in België. Het rapport bevat o.a.:

- een beschrijving hoe de oprichting van het data-archief juridisch moet worden geregeld; welke garanties er moeten worden geboden
- welke technische vereisten zijn en de financiële implicaties zijn
- een beschrijving van de kritische succesfactoren , de stakeholders en de partners
- formulieren, modelcontracten, model van metadataseten

#### Duurtijd van het onderzoek

Het onderzoek moest in een periode van 6 maanden worden uitgevoerd en afgesloten door een eindrapport.



### Begeleidingscomité

Het onderzoeksteam heeft zich laten bijstaan door een begeleidingscomité. Dit begeleidingscomité was samengesteld uit externe personen die de nodige bekwaamheid bezitten in de betrokken wetenschappelijke, technische en/of functionele domeinen.

De diensten van de POD Wetenschapsbeleid hebben beslist over de uiteindelijke samenstelling van het begeleidingscomité, en heeft de vergaderingen van het comité voorgezeten.

Tussentijds heeft ook een technisch begeleidingscomité vergaderd om over bepaalde, technische onderwerpen, in de diepte te vergaderen.

### Seminarie

Voor de afwerking van het eindrapport, en ter toetsing van de voorlopige, eerste resultaten werd een internationaal seminarie ingericht. Het programma van dit seminarie, kan gelezen worden in Bijlage/Annexe 8

### Het onderzoeksteam

De opdracht werd toevertrouwd aan drie onderzoeksinstellingen die elk een deel van het onderzoek voerden, in functie van de eigen specialisatie van de betrokken onderzoeksinstelling:

1. **Cellule Interfacultaire de Technolgy Assessment, van de (FUNDP) Facultés universitaires Notre-Dame de la Paix à Namur / *The Technology Assessment Research Unit from the University of Namur (CITA)***  
<http://www.fundp.ac.be/cita><sup>2</sup>

Onder leiding van Prof. Claire LOBET-MARIS; onderzoekster Véronique LAURENT. Dit onderzoekscentrum is gespecialiseerd in het voeren van assessments van nieuwe technologieën, in het bijzonder inzake informatie- en communicatie-technologie (ICT). In dit onderzoeksproject onderzoekt dit team de inhoudelijke taken van het data-archief, het onderzoek van de noden en behoeften van de onderzoekers en gebruikers, zowel inzake deponering van data als inzake later gebruik van de gearchiveerde data.

2. **Centre de Recherche Informatique et Droit van de (FUNDP) Facultés universitaires Notre-Dame de la Paix à Namur / *The Centre for Computer and Law from the University of Namur (CRID)***  
<http://www.crid.be><sup>3</sup>

Onder leiding van Prof. Yves POULLET; onderzoekster Nathalie LEFEVER. Dit onderzoekscentrum legt zich toe op de studie van juridische problemen en casussen bij ICT. In dit onderzoek leggen de onderzoekers zich toe op alle juridische vragen die zich stellen bij de inrichting van een dergelijk data-archief, zoals daar zijn : data protection, bescherming van auteurs- en andere intellectuele rechten, institutionele bezwaren, gebruik van de data door derden.

---

<sup>2</sup> Prof. Claire Lobet-Maris, Véronique Laurent, Cellule Interdisciplinaire de Technology Assesement FUNDP, [clo@info.fundp.ac.be](mailto:clo@info.fundp.ac.be), [vla@info.fundp.ac.be](mailto:vla@info.fundp.ac.be)

<sup>3</sup> Prof. Yves Poulet, Nathalie Lefever, Centre de Recherche Informatique et Droit FUNDP, [yves.poulet@fundp.ac.be](mailto:yves.poulet@fundp.ac.be), [nathalie.lefever@fundp.ac.be](mailto:nathalie.lefever@fundp.ac.be)

3. **The Expertisecentrum David (eDAVID), Antwerpen**

<http://www.expertisecentrumdavid.be><sup>4</sup>

Vertegenwoordigd door Inge SCHOUPS.

Het Expertisecentrum DAVID is in Vlaanderen het onderzoekscentrum inzake digitale archivering en duurzame bewaring van digitale data. In dit onderzoek staat dit centrum voor een stand van zaken van de technische archiveringskwesties en technieken nodig voor de archivering van de data en bijhorende documenten.

---

<sup>4</sup> Inge Schoups, Expertisecentrum DAVID, [inge@edavid.be](mailto:inge@edavid.be)





## 1ERE PARTIE : ANALYSE DE L'EXISTANT

Avant de pouvoir faire un diagnostic de faisabilité d'un service d'archivage pour les sciences sociales en Belgique, il était très important d'étudier les expériences déjà réalisées en la matière. Cette première partie du rapport fait donc le point sur la question, dans un premier temps en Belgique, ensuite à l'étranger, et enfin, dans le contexte européen. Cette phase "d'inventaire" fut essentielle pour nous permettre de proposer un projet de service d'archivage bien adapté au contexte particulier de la Belgique.

### Introduction : Méthodologie

#### 1. PANORAMA DES SERVICES D'ARCHIVAGE

Après un premier tour d'horizon de différents portails et services d'archive en sciences sociales disponibles en annexe 1 de ce rapport, nous avons conclu qu'il était très difficile de tirer des généralités de nos observations. En effet, les services se différencient très fortement par leur taille, leur développement en termes de services d'archivage, leur organisation, les types de données qu'ils proposent, ...

Nous avons donc choisi de porter notre attention sur une étude en profondeur de certains services et portails d'archivage, à la fois différents et complémentaires, que nous avons 'sélectionnés' pour les raisons ci-dessous.

- En Belgique : le "Belgian Archives for the Social Sciences" (BASS), à Louvain-La-Neuve : <http://bass.rspo.ucl.ac.be/>

*Pour réaliser une étude de faisabilité, nous nous devons de faire un état des lieux de l'existant belge.*

- En Allemagne : le "Zentralarchiv für Empirische Sozialforschung", archives centrales de la recherche sociale empirique (ZA/GESIS), à Cologne :  
<http://www.gesis.org/en/za/index.htm>

*Ce service d'archivage a été choisi pour sa spécificité axée sur la recherche sociale empirique. Il est également l'un des centres les plus développés en Europe. Le ZA fait partie du GESIS, une institution de services de recherche membre du consortium de recherche Gottfried Wilhelm Leibniz, recevant actuellement des fonds de l'Etat fédéral et des fonds des Etats allemands.*

- En Grande-Bretagne : l'UK Data Archive (UKDA), à Colchester :  
<http://www.data-archive.ac.uk>

*Ce service d'archivage est également très développé. Il est le plus grand conservateur de données digitales concernant les sciences humaines et le genre humain en UK. Nous l'avons choisi pour son portail très complet, donnant accès aux procédures de publication et d'accès aux données dans une grande transparence. Contrairement au centre allemand, il présente aussi la particularité de faire partie d'une université, option qu'il était intéressant d'étudier.*

- En Europe: le Council of European Social Science Data Archives (CESSDA) :  
<http://www.nsd.uib.no/cessda/>

*Au niveau européen, il nous est paru intéressant de marquer un arrêt sur ce portail 'parapluie' qui coordonne les efforts de centres nationaux d'archive de données en sciences sociales.*

## 2. LA COLLECTE DE DONNEES

Pour analyser l'existant dans le domaine de l'archivage en sciences sociales, la collecte de nos données s'est réalisée en trois temps.

Dans un premier temps, nous avons réalisé une **desk-analysis**, c'est-à-dire une analyse des sites Internet des portails d'archive en sciences sociales ciblés ci-dessous. Cette analyse très utile a pu nous donner de nombreux indices sous les angles sociologiques et juridiques. Nous avons ainsi pu recueillir plusieurs informations sur les services proposés par les centres d'archivage, les types de données et de ressources proposées, leur organisation générale et leur statut juridique, les conditions et les modes d'accès aux portails, les modes de gestion des sites Internet d'archives, ... Les grilles d'analyses que nous avons employées sont disponibles en annexe 2 et en annexe 3 de ce rapport.

Dans un deuxième temps, nous avons complété nos observations par des **visites** et des **interviews** avec plusieurs responsables de centres d'archives. Ces visites nous ont permis d'avoir une vision beaucoup plus complète des missions et de l'organisation pratique des services d'archivage. Un exemple de questionnaire-type que nous avons soumis aux responsables rencontrés se trouve en annexe 4 de ce rapport.

Les interviews et les visites suivantes ont donc été réalisées :

- Une visite du ZA-GESIS et une rencontre avec plusieurs représentants le 19 juin 2007
- Une visite de l'UK Data Archive et une rencontre avec Kevin Schürer, également directeur du CESSDA, le 12 juillet 2007
- Une interview avec un responsable du BASS le 20 juillet 2007

Enfin, dans un troisième temps, un **séminaire international** organisé le 16 octobre 2007 à Bruxelles, en compagnie d'experts internationaux venant notamment d'Allemagne, d'Angleterre et de Suède, nous a permis d'affiner notre perception de l'existant dans le secteur de l'archivage en sciences sociales.

### 3. ANALYSE DES DONNEES

Toutes les informations que nous avons collectées ont été retranscrites et analysées selon des grilles standardisées. Nous vous présentons ci-dessous les résultats de ce travail. L'obtention de toutes ces données n'aurait pas pu être possible sans la collaboration très précieuse des nombreuses personnes que nous avons rencontrées et interviewées. Qu'elles en soient, une fois de plus, chaleureusement remerciées !

Notons que les services que nous avons analysés en profondeur ne sont que des cas différents mais non représentatifs de la variété des situations du domaine de l'archivage. Une étude représentative de l'archivage en sciences sociales nécessiterait une enquête chiffrée de grande envergure. Notre temps étant limité dans le cadre de cette étude, nous nous sommes uniquement consacrés à quelques cas.

## Chapitre 1 : Archivage en Belgique

---

### 1.1. BELGIAN ARCHIVES FOR THE SOCIAL SCIENCES (BASS)

Université catholique de Louvain  
**B.A.S.S. Belgian Archives for the Social Sciences**  
**Archives Belges en Sciences Sociales**  
*Bâtiment Leclercq*  
place Montesquieu, 1 bte 18  
B-1348 Louvain-la-Neuve (Belgique)  
<http://bass.rspo.ucl.ac.be/>

De doelstellingen van BASS zijn :

1. de verwerving, bewerking, ter beschikkingstelling en archivering van data, in de onderzoeksdomeinen demografie, economie en politieke wetenschappen
2. verwerken van alle vragen van onderzoekers van de gehele faculteit

Het centrum was en is sedert 1970 verbonden aan de Université Catholique de Louvain (UCL) te Louvain-La-Neuve, en was met name een initiatief van het departement sociologie waar het nog steeds aan verbonden is. De medewerkers van BASS zijn informatici en een socioloog. In de eerste jaren van de werking tot de periode 1985 werkten er 5 à 8 medewerkers. Het beheer van de data-archieven is thans doorgegeven aan de technische en logistieke dienst van faculteit waaraan een aantal informatici verbonden zijn; het vroegere universitaire rekencentrum is verzelfstandigt. Het IDQL (Installation facultaire LECLERCQ) is een logistieke dienst voor alle vestigingen in de onmiddellijk omgeving van het collège Jacques LECLERCQ.

Sedert september 1998, heeft het IDLQ de ondersteuning van BASS: Belgian Archives in social Sciences overgenomen en maakt BASS dus minder deel uit van een onderzoeksafdeling.

BASS sloot aan bij de startende initiatieven in Europa inzake het verzamelen en beheren van sociale data, en was op deze wijze een van de eerste centra voor de eurobarometers. In verband met de Eurobarometer, was België met BASS in 1970 één van de vijf initiatiefnemers voor het European Communities Study, en zo de start van de tweejaarlijkse Eurobarometers. BASS heeft in het verleden in de Europese gemeenschap een belangrijk aandeel gehad in de ontwikkeling en vastleggen van codeboeken en de eerste ontwerpen van metadatasets. De oprichting in 1977 (22 mei 1977) van de International Federation of Data Organizations (IFDO), een van de belangrijke spelers om een internationaal netwerk te vormen, ging door in Louvain-La-Neuve. Vandaag vertegenwoordigen de IFDO-leden een netwerk van meer dan 30 data-archieven wereldwijd.

De werking heeft zich later uitgebreid voor data-archivering voor diverse Belgische universiteiten zodat BASS ook bestanden beheert van andere universiteiten dat de eigen.

Momenteel bestaat BASS nog steeds maar wordt er niet actief wervend opgetreden inzake data-archieven en op zoek gegaan of gesensibiliseerd om data te deponeren bij BASS. De rol kan best omschreven worden als die van een 'relais' of postbus; er zijn de laatste jaren nog data deponeringen gebeurt. Dankzij de goede banden met andere Europese data-archieven worden bestanden doorverwezen of ter beschikking gesteld. Zo worden data doorverwezen naar ZA (Duitsland) of Steinmetz/DANS (Nederland).

Niettegenstaande de minder actieve werking, rol en aanwezigheid, blijft BASS vragen krijgen van buitenlandse onderzoekers naar gearchiveerde 'Belgische data'.

De reeks gearchiveerde enquêtes betreft de periode 1972 tot 1998. De website is echter na 1998 niet meer geactualiseerd.

Wat de Europese data betreft, blijft BASS zijn rol spelen. Eén medewerker blijft een opdracht hebben inzake verspreiding van eurostat/eurobarometer en blijft de maandelijkse aanvullingen verzorgen.

België was in de zin een van de vroegste leden van het Europese netwerk. BASS is steeds aangesloten gebleven bij CESSDA<sup>1</sup>, en bij de leden van het Europese netwerk bestaat een grote goodwill om het Belgisch lidmaatschap te behouden. De stille Belgische vertegenwoordiging van de laatste jaren wordt weliswaar betreurd.

Belangrijke data gegenereerd op Europees niveau, zoals de European Election Studies die door de Europese instanties worden gedeponereerd bij ZA (Duitsland), worden door het ZA doorgestuurd naar BASS.

De huidige databeheerders van BASS menen dat de daling van de aangeboden data-archivering ook externe oorzaken heeft.

- Zo zijn de beschikbare middelen voor onderzoek, zeker in Wallonië gedaald, ten voordele van Europese onderzoeksprojecten en -gelden. De resultaten van deze onderzoeken volgen de weg van archivering van andere Europese data, naar het ZA.
- Ook instellingen zoals FWO, POD Belspo of Nationale Lotterij spelen een steeds grotere rol voor de financiering van bijzondere onderzoeksprojecten. Met deze instellingen zijn geen afspraken gemaakt over de verdere bewaring of deponering van de resultaten van hun onderzoeksopdrachten.
- Grote onderzoeken waarvoor veel statistisch materiaal moet worden verzameld of verwerkt zijn lang niet meer het werk van of te bekostigen door één onderzoeksgroep binnen één universiteit maar een zaak van samenwerking tussen universiteiten met een groot aandeel van andere instellingen zoals het INS of FODs.

---

<sup>1</sup> Zie hierna over de werking van CESSDA

- De focus aan de universiteiten is verschoven naar de onderwijsopdracht, en minder bij onderzoekstaken. Voor onderzoek, en zeker voor langlopende projecten moeten steeds vaker externe middelen worden gezocht en verlopen deze met meerdere partners. Hierbij worden zelden afspraken gemaakt over waar de data na verloop worden bewaard.

#### Toekomstpunten

- Binnen het CESSDA-netwerk zal het belangrijk zijn dat er één Belgische lidmaatschap is.
- Er is een belangrijke rol weggelegd voor de FNRS-/FWO-instellingen in beide landsdelen die veel meer dan de universiteiten promotor zijn van grote onderzoeksprojecten
- Uit de gesprekken lijken de BASS-medewerkers aan te geven dat een federaal archief-initiatief een groter impact zou hebben
- Is er een rol voor een instellingen zoals de Nationale Loterij
- BASS ziet nog steeds een rol voor zichzelf in de Franstalige gemeenschap, en niet meteen een rol binnen de federale, Belgische context
- Wanneer een universiteit (opnieuw) de rol van een data-archief zou opnemen, moet de structuur zo worden georganiseerd dat belangrijke instanties van de overheden die grote data beheren, vertegenwoordigd zijn.
- Belang van verdrag van Berlijn.

## 1.2. HET RIJKSARCHIEF

Algemeen Rijksarchief en Rijksarchief in de provinciën<sup>2</sup>

Ruisbroekstraat 2

1000 Brussel

Tel: (+32) 02/513.76.80

<http://arch.arch.be/>

Karel Velle, Algemeen Rijksarchivaris

Het Algemeen Rijksarchief en Rijksarchief in de provinciën is een wetenschappelijke instelling van de Belgische Federale Overheid, die deel uitmaakt van de POD Wetenschapsbeleid. De instelling bestaat uit het Algemeen Rijksarchief te Brussel en uit de 19 Rijksarchieven verspreid over het gehele land.

Het Rijksarchief oefent toezicht uit op de goede bewaring van de archieven die door een overheid werden tot stand gebracht en beheerd. Het geeft in dit verband richtlijnen en aanbevelingen, verricht controlebezoeken, organiseert cursussen voor ambtenaren en treedt op als raadgever voor de bouw en de inrichting van archiefbewaarplaatsen en voor de organisatie van het archiefbeheer binnen een overheidsbestuur. Archieven kunnen na minstens 30 jaar worden overgedragen.

Databestanden die worden aangemaakt door overheidsdiensten, bij de uitvoering van hun taak, maken deel uit van het geheel van de archieven van die betreffende overheidsdienst en worden overeenkomstig de bepalingen van de Archiefwet, overgedragen aan het Rijksarchief. Bij deze archiveringsplicht wordt geen onderscheid gemaakt in vorm of drager; aan het Rijksarchief worden dan ook heel wat digitale bestanden aangeboden ter archivering. In die zin bewaart het Rijksarchief heel wat data uit enquêtes en statische verzamelingen. Deze data worden bewaart, bewerkt en voor onderzoekers ter beschikking gesteld.

---

<sup>2</sup> Zie ook Deel drie.



Behalve de archivering van overheidsbestanden en –documenten, kan het Rijksarchief ook archieven van privé oorsprong, archieven van personen, verenigingen, bedrijven die een belangrijke taak of rol in de Belgische gemeenschap vervullen opnemen, beheren en ter beschikking stellen.

De Belgische situatie is zeer vergelijkbaar met die van de andere Europese landen waar naast een nationale archiefdienst een data-archief voor onderzoekers bestaat. De werking van BASS en het Rijksarchief hebben in het verleden zonder problemen bestaan.

### 1.3. DISSCO

Een onderzoeksproject waarin de POD Wetenschapsbeleid evenzeer sponsor is.  
Website : <http://www.meteo.be/dissco>

Het onderzoek wordt gevoerd aan  
Vrije Universiteit Brussel (VUB)  
Opleiding Archivistiek en Hedendaags Documentbeheer (OAHD)  
Pleinlaan 2  
1050 BRUSSEL  
Tel: 02 629 24 34  
fscheeli@vub.ac.be

Dit onderzoeksproject loopt van 2001 tot 2008. Het dient hier kort te worden vermeld omdat het zich richt tot betere ondersteuning en documentering van het proces van wetenschappelijk onderzoek.

DISSCO ontwikkelt een records management applicatie ter ondersteuning van administratieve basisprocessen in wetenschappelijke instellingen. De klemtoon ligt bij de ontwikkeling van functionaliteiten om de documentenstroom en –beheer te ondersteunen, en voor het beheer van statisch archief. Het DISSCO-systeem gaat uit van het werken met standaarden en open source.

De gebruikersdoelgroep bestaat uit wetenschappelijke instellingen (van de Staat of andere instellingen zonder winstoogmerk die een wetenschappelijke opdracht vervullen, b.v. universiteiten). De applicatie die tijdens dit project werd ontwikkeld, werd bij de projectpartners getest: het Koninklijk Meteorologisch Instituut van België (KMI), het Studie- en Documentatiecentrum Oorlog en Hedendaagse Maatschappij (SOMA), de Université Libre de Bruxelles (ULB) en de Vrije Universiteit Brussel (VUB). In principe moet de applicatie in een groot aantal wetenschappelijke instellingen bruikbaar zijn.

De doelstellingen van DISSCO zijn breder dan de archivering van de data, resultaat van het wetenschappelijk onderzoek. De doelstellingen liggen in de ondersteuning van het documentbeheer en de informatiestroom in wetenschappelijke processen en activiteiten. Daartoe heeft het project onderzocht in welke mate de bestaande standaarden en modellen voor digitaal document- en archiefbeheer, kunnen worden toegepast. Het gaat dan om hoofdzakelijk om: ISO 15489, *Model Requirements for Electronic Records Management Systems*, en ISAD(G), *General International Standard Archival Description*.

Het project heeft onderzoek gevoerd naar welke software deze document- en informatiestroom best kan ondersteunen. Voor het onderzoek naar de haalbaarheid van een data-archief, is het

DISSCO niet zonder belang. Een goede archivering van de data van wetenschappelijk onderzoek is slechts mogelijk wanneer ook alle documentatie over het verloop van het onderzoek, goed wordt vastgelegd en gearchiveerd. In de discussie over de metadata van de gearchiveerde data.

## **Chapitre 2 : Analyse des expériences étrangères**

---

### **2.1. L'ALLEMAGNE : L'EXPERIENCE DU ZA/GESIS**

#### **2.1.1. Création et évolution**

Les archives centrales de la recherche sociale empirique ou Zentralarchiv für Empirische Sozialforschung (ZA : <http://www.gesis.org/en/za/index.htm>) ont été créés en 1960 sous la forme d'un Institut de la Faculté d'Economie et des Sciences Sociales de l'Université de Cologne.

A partir de 1986, elles ont pris la forme d'une association, et en 1996, sont devenues une entité constitutive du GESIS, le Gesellschaft Sozialwissenschaftlicher Infrastruktureinrichtungen ou German Social Science Infrastructure Services.

Le GESIS fournit des services pour soutenir la recherche en sciences sociales comme le développement, la création ou l'archivage de bases de données de recherche ou de littérature en sciences sociales. Il mène aussi des réflexions méthodologiques et de développer des méthodes de recherche sociale empirique. Le GESIS combine d'ailleurs fortement des activités d'archivage avec des activités de recherche.

Le GESIS est divisé en **trois centres locaux**

- Le **Social science information centre (IZ)**, basé à Bonn. Ce centre s'occupe de la collecte et de la diffusion d'information sur l'état de la recherche et de la littérature dans les sciences sociales dans les pays de langue allemande et dans les autres pays de l'Europe de l'Est. Le centre s'occupe aussi de faire de la recherche et développement dans les sciences et les technologies de l'information (<http://www.gesis.org/en/iz/index.htm>).
- Le **Centre for Survey Research and Methodology (ZUMA)**, basé à Mannheim. Ce centre offre des services de consultation pour établir, conduire et analyser des enquêtes de sciences sociales destinées aux personnes travaillant dans la recherche sociale. Il fait également de la recherche pour améliorer les bases méthodologiques et techniques de la recherche en sciences sociales (<http://www.gesis.org/en/zuma/index.htm>).
- Le **Central Archive for Empirical Social Research (ZA)**, basé à Cologne et collaborant de manière très proche avec l'Université. Ce centre s'occupe de l'archivage de données primaires (données, questionnaires, variables) et de données sociales empiriques pour les préparer à des analyses secondaires et les rendre accessibles au public. Il est aussi expert dans les procédures techniques de la recherche sociale empirique et historique.

En plus de ces trois services, le GESIS possède une agence de service centrale pour l'est de l'Europe (Gesis Service Agency Eastern Europe : [http://www.gesis.org/en/gesis\\_branch\\_office/index.htm](http://www.gesis.org/en/gesis_branch_office/index.htm)). Cette agence est spécialisée dans le transfert d'information et de données en sciences sociales entre l'est et l'ouest de l'Europe.

### **2.1.2 Missions, organisation et financement**

La **mission du ZA** est essentiellement de créer une infrastructure optimale pour la recherche sociale allemande et internationale. Celle-ci comporte plusieurs points :

- permettre l'exploitation secondaire de données
- créer un réseau d'archive
- mettre sur pieds des programmes d'entraînement et d'apprentissage
- publier des données
- faire de la recherche en sciences sociales

Le **portail d'archive** développé par le service sert de stockage et de relais entre les déposants et les utilisateurs de données. Ses missions comprennent :

- la collecte des données auprès de départements gouvernementaux allemands (sondages électoraux depuis 1949), d'études comparatives européennes (Eurobaromètre, ISSP), de recherches privées (ALLBUS), d'instituts d'enseignements et de recherches ou de chercheurs indépendants ;
- la vérification de la qualité des données et leur préparation pour la réutilisation ;
- la réalisation d'un répertoire des données disponibles (le ZACAT) ;
- la conservation des données et leur stockage ;
- la mise à disposition de ces données pour une réutilisation, dans le respect des souhaits de l'auteur de la collecte originale et via l'octroi de droits d'accès ;
- l'apport d'un support et de conseils pour le traitement des données en vue d'une réutilisation ;
- l'échange de données avec d'autres pays germanophones ou avec l'étranger via des organisations internationales ;
- la recherche dans le domaine de la conservation des données et de leur analyse ;
- la rédaction d'une publication bi-annuelle sur les tendances en recherche sociale ;
- l'organisation de conférences et la rédaction de conventions.

#### **A. ORGANISATION**

Depuis 1960, l'organisation n'a cessé de s'agrandir pour s'adapter à la demande des utilisateurs et du CESSDA. Le ZA compte donc aujourd'hui environ **50 travailleurs équivalents temps-plein**, c'est à dire 42 employés et environ 8 étudiants. Il est divisé en plusieurs départements : acquisition, vérification des données, valorisation des données, documentation, accès aux données, administration informatique et relations publiques. Les employés sont généralement spécialisés par départements et par types de données (nationales, européennes et internationales).

#### **B. FINANCEMENT**

Le ZA fait partie du GESIS, une institution de services de recherche membre du consortium de recherche Gottfried Wilhelm Leibniz. En tant qu'institution de services de recherche, le GESIS reçoit actuellement **des fonds de l'Etat fédéral** (pour 80%) et des **fonds des Etats**

**allemands pour 20%.** Parmi ces 20% des fonds en provenance des Etats allemands, 25% des fonds sont octroyés par les Etats où se situent les trois centres d'archivage (Cologne, Bonn et Mannheim) (75% au total) et 25% des fonds sont octroyés par tous les autres Etats.

### 2.1.3. Types de données

#### C. CARACTERISTIQUES DES DONNEES

Depuis sa création, la banque de données du ZA a considérablement augmenté, passant de 9 études à environ 5000 aujourd'hui.

Les données conservées, accessibles gratuitement, sont largement **quantitatives** et issues de **différentes disciplines des sciences sociales** telles que la sociologie, l'histoire, la criminologie, la démographie, la statistique<sup>3</sup>, ...

Le ZA collecte des données provenant de **bases de données privées ou publiques** venant de nombreux donateurs nationaux ou internationaux. Parmi ceux-ci, on peut trouver :

- ALLBUS : données sociales générales concernant l'Allemagne
- Eurobarometer : études européennes
- International Social Survey Programme (ISSP)
- D'autres collections de données spéciales

Dans le portail d'archive, ces données sont **regroupées par catalogues** ou par bases de données à partir desquels les utilisateurs peuvent consulter de la documentation, télécharger et réutiliser des données. Pour chaque regroupement, des conditions spécifiques d'accès aux données sont d'application (nécessité de s'enregistrer ou non pour accéder au portail, accès possible à des bases de données ou uniquement à de la documentation, ...).

#### Data - Retrieval & Downloads<sup>4</sup>

- [ZA Data Holding Catalogue](#) - Search the study descriptions of the complete data holding (catalogue principal et général)
- [ZA - Qbase](#) - Search codebooks and original questionnaires of selected data collections
- [ZACAT](#) - ZA Online Study Catalogue - Browse, analyse & download survey data like ISSP, Eurobarometer, EVS, on German elections, from Eastern Europe, on Youth & Adults (registration free of charge)
- [ZA Codebook Explorer Internet version](#) - Browse studies or trends, search for keyword or analyse and download data on studies from selected databases (registration free of charge)
- [ALLBUS Compact](#) Download for scientific research, training & teaching (German General Social Survey) (registration free of charge)
- [ALLBUS full version](#) - Download for personal scientific research (registration free of charge)
- [HISTAT DB](#) - Historical Statistics Online (in German - registration free of charge)
- [ZA Library 'Empirische Sozialforschung'](#) - e.g. literature related to archived studies
- [IFDONet](#) (external link) - International Federation of Data Organizations for the Social Science - Access to data sources of the international archive network

<sup>3</sup> Notons que la banque de données du ZA compte très peu de références en économie.

<sup>4</sup> Extrait de la page web : [http://www.gesis.org/en/data\\_service/index.htm](http://www.gesis.org/en/data_service/index.htm)

## D. MODE DE SELECTION DES DONNEES

La philosophie du ZA est de ne pas faire de sélection en faveur de certaines disciplines de données, mais d' **intégrer toutes les disciplines** dans un but d' **interopérabilité**.

Malheureusement, par manque de moyens pour pouvoir conserver toutes les données, le service doit opérer une certaine sélection ! Cette sélection s'adapte généralement à la demande des utilisateurs et du plus grand nombre.

.Les données sont généralement conservées par le ZA à la condition que les études :

- soient issues des sciences sociales ;
- s'appliquent à la population allemande ou à une partie de celle-ci;
- aient été réalisées par des chercheurs allemands, qu'elles soient relatives à l'Allemagne ou non;
- soient d'envergure nationale ou internationale;
- soient représentatives d'une population.

Les données locales, en revanche, s'avèrent plus rarement conservées.

### ***2.1.4 Collecte et mise à disposition des données***

## E. PROCESSUS DE COLLECTE ET DE MISE A DISPOSITION DES DONNEES

La collecte de données et l'alimentation du portail du ZA s'effectue en plusieurs étapes.

### **1) Acquisition des données :**

- Contacts avec les fournisseurs de données, les instituts de recherche, les organisations publiques, ...
- Acquisition de publications pertinentes et de rapports scientifiques.

### **2) Transformation et nettoyage des données collectées (enlèvement des erreurs...)**

- Uniformisation des données à l'aide d'outils pour nettoyer les données, Ces outils sont particulièrement employés pour les études internationales.

### **3) Méta-documentation :** description des données.

### **4) Préservation des données et des documents.**

- Archivage digital des données.
- Classement de celles-ci de manière à pouvoir les retrouver facilement dans un portail informatisé.
- Adaptation aux changements technologiques.

### **5) Mise à disposition des données:**

- Production de catalogues de documentation, mise à disposition des données sous forme de CD ou de téléchargement.
- Un certain nombre de **services** pour la consultation des données sont disponibles sur la plateforme du ZA comme de la documentation multilingue pour permettre des

comparaisons internationales, ou des outils d'analyse, de visualisation ou de distribution des données<sup>5</sup>.

## 6) Promotion des données

- Le ZA peut compter sur un **support en réseau** pour promouvoir ses données. Le CESSDA, l'Europe Centrale et de l'Est, l'European data laboratory, et les programmes d'enseignement et d'entraînement organisés à l'Université de Cologne ou dans d'autres universités (séminaires, workshops, journées d'études, conférences, cours donnés sur l'analyse des données, ...) peuvent jouer ce rôle de support en réseau.
- Enfin, la maison compte un département de **relations publiques**, chargé d'éditer des journaux du ZA et du GESIS (ZA-information), présents sur Internet et certaines publications.

## F. TECHNIQUES D'ACQUISITION DES DONNEES

**D'après les travailleurs du ZA que nous avons rencontrés, l'acquisition de données prends généralement un certain temps.**

**Différents moyens peuvent être utilisés pour convaincre des fournisseurs à partager leurs données :**

- 1) Envoyer des questionnaires à des personnes ayant mené des recherches.
- 2) Inviter des jeunes chercheurs à des séminaires sur l'analyse de données pour les sensibiliser au partage de données.
- 3) Profiter de programmes de collecte de données comme ALLBUS, eurobarometer, ...
- 4) Employer des exemples liés à des disparitions historiques de données pour convaincre des personnes de les partager.

L'acquisition des données nécessite également qu'un bon contact ou de bonnes négociations soient entreprises entre les gestionnaires du service d'archives et les fournisseurs de données. Il est donc nécessaire de créer une **situation « gagnant-gagnant »** entre les deux parties.

**Différents** moyens peuvent être employés par des gestionnaires du service d'archives pour convaincre des dépositaires de données de confier celles-ci à un portail d'archive<sup>6</sup>, comme ceux d'offrir aux dépositaires différentes opportunités :

- Obtenir une conservation sécurisée de leurs données, en cas de perte.
- Avoir un accès totalement gratuit à d'autres types d'études en devenant dépositaires dans le portail (« prime customers »).
- Rendre les données en sciences sociales utiles à la communauté.
- Diffuser les résultats de recherche dans le monde entier.

### 2.1.5 Dépôt, accès et réutilisation des données

La philosophie du ZA est en faveur d'un accès le plus ouvert possible aux données.

---

<sup>5</sup> Le portail ZA fait ainsi un lien vers le système NESSTAR d'encodage et d'analyse de données : <http://www.nesstar.com/>

<sup>6</sup> Nous développerons plus en profondeur ces aspects du dépôt de données par les fournisseurs dans la troisième partie de ce rapport consacrée à la mise en place du service d'archivage.

## G. DEPOT DE DONNEES

La fixation de l'accès aux documents dans le portail du ZA est confiée aux **dépositaires** de données. Chaque type de fournisseur de données dans le portail ZA peut donc fixer des conditions d'accès à ses propres données en choisissant parmi quatre catégories proposées.

- Catégorie 0 : Données et documents accessibles à tout le monde.
- Catégorie A: Données et documents réservés à la recherche académique et à l'enseignement.
- Catégorie B : Données et documents réservés à la recherche académique et à l'enseignement *à condition que ces documents ne soient pas publiés* (si ces données veulent être publiées ou travaillées la ZA doit donner une permission).
- Catégorie C : Données et documents réservés à la recherche académique et à l'enseignement *après que le dépositaire en aie donné son autorisation écrite* (permission accordée d'après les précisions de chaque utilisateur sur leurs intentions d'analyse).

Notons que cet **accès** aux ressources peut être **total ou partiel**.

Les dépositaires du portail ne signent pas de contrat de dépôt de leurs bases de données dans le ZA. Ils doivent prendre connaissance de la charte d'utilisation des données<sup>7</sup> et ont donc la liberté de fixer des catégories d'accès à leurs données. C'est bien souvent un accès le plus large possible aux données qui est fixé, sauf dans le cas de données sensibles (données pénales ou données portant sur le marché du travail par exemple), pour lesquelles des conditions spéciales d'accès s'appliquent généralement.

## H. ACCES ET COMMANDE DES DONNEES

Comme nous l'avons montré ci-dessus, chaque type de catalogue dans le portail du ZA requière différentes conditions d'accès aux données (nécessité d'un enregistrement ou non dans le système, accès à des fichiers de données ou uniquement à un résumé de la recherche...).

Si toutes les données ont généralement des fiches descriptives accessibles en ligne, les utilisateurs du portail doivent fréquemment passer par un **service de commande** (ZA Order Service) pour obtenir les fichiers (datasets) ou la documentation additionnelle. Les utilisateurs devront, pour se faire, remplir des formulaires ou des contrats de commande accessibles en ligne.

Ces formulaires ou contrats de commande sont généralement spécifiques à chaque type de données : studies in general, Allbus, election studies, ISSP, Eurobarometer, Eastern Europe,... Cependant, certains types de données sont directement téléchargeables et ne nécessitent pas le remplissage d'un formulaire. C'est le cas pour les questionnaires et des études comme les « election studies ».

Pour commander des données en ligne, l'utilisateur doit remplir un formulaire où il devra marquer son accord avec les procédures de réutilisation des données ("ZA regulations of use", voir ci-dessous).

### Dans ce formulaire, il devra ensuite

1) Cocher les supports de données qu'il souhaite obtenir.

---

<sup>7</sup> ZA regulations of use : [http://www.gesis.org/en/data\\_service/order/usage\\_regulations.htm](http://www.gesis.org/en/data_service/order/usage_regulations.htm)  
Nous présenterons plus bas cette charte d'utilisation.

Le matériel est en effet habituellement livré sur cd-rom, un support payant.

2) Commander le matériel qu'il souhaite obtenir :  
(codebook en format pdf, fichiers SPSS en format sav, rapports méthodologiques en format pdf)

3) Décrire brièvement quelle utilisation il souhaite faire du matériel qu'il demande.

4) Remplir ses coordonnées (nom, prénom, adresse, e-mail, téléphone) pour chaque ressource qu'il souhaite commander (il n'existe pas d'enregistrement général au système permettant d'accéder à toutes les ressources).

## I. PROCEDURES DE REUTILISATION DES DONNEES PAR DES TIERS

Pour pouvoir employer les données du portail, les utilisateurs du service d'archivage doivent adhérer à un certain nombre de réglementations sur la réutilisation des données, c'est-à-dire aux "ZA regulations for use"<sup>8</sup>. Ces réglementations sont les suivantes:

- Les documents disponibles sur le portail sont valables uniquement pour des analyses scientifiques dans le cadre de la recherche et de l'enseignement.
- Les dépositaires des données fixent les conditions d'accès à leurs documents.
- Un réutilisateur de données doit
  - informer la plateforme quand la recherche réutilisant les données est terminée ;
  - ne pas transmettre les données à un tiers ;
  - effacer les données dès que la recherche est terminée ;
  - citer la recherche originale dans la nouvelle recherche ;
  - envoyer deux copies des sources des publications sur lesquelles il s'appuie pour la publication au ZA ;
  - payer pour l'obtention des ressources (les ressources sont gratuites si elles sont téléchargées directement en ligne mais payantes pour les obtenir sur cd).

Remarquons que le site Internet encourage les utilisateurs ne vivant pas en Allemagne à faire directement leurs demandes dans leur propre pays. La nécessité d'un service d'archivage en Belgique se justifie donc d'autant plus !

## J. PRIX DES DONNEES

Différents critères rentrent en ligne de compte pour calculer les prix des documents mis à disposition<sup>9</sup> :

- La **manutention nécessaire** et les **coûts de licence** pour produire le support. (Produire et envoyer un fichier sur ftp coûte par exemple moins cher que de le convertir en cd-rom).
- Le **statut de l'utilisateur** (les personnes et les instituts fournissant du matériel ou des études à la ZA peuvent y accéder gratuitement, s'il n'y a pas de restrictions d'accès).
- Les **types de CD-rom** (si ils sont « standard », avec des outils, « customisés », ...).

Les **prix des ressources (au 22 juin 2006)** sont les suivants :

	Academic and	Non-academic	Students
--	--------------	--------------	----------

<sup>8</sup> ZA regulations of use : [http://www.gesis.org/en/data\\_service/order/usage\\_regulations.htm](http://www.gesis.org/en/data_service/order/usage_regulations.htm)

<sup>9</sup> Pour un aperçu détaillé des prix des ressources, vous pouvez consulter le site [http://www.gesis.org/en/data\\_service/order/charges.htm](http://www.gesis.org/en/data_service/order/charges.htm)



	research training	user	
<b>ZA CD Rom standard</b> (Standard editions, e.g. all Allbus-studies, all Politbarometer-studies)	25€	50€	25€
<b>ZA CD Rom plus</b> Added-value editions with extra tools (ex JD-system, ZA codebookexplorer,...) for collections	50€	100€	50€
<b>ZA CD Rom - customized</b> Data and documentation customized to user needs <u>and delivered on CD-Rom</u>			
➤ Up to 5 studies	30€	60€	20€
➤ 6-10 studies	50€	100€	20€
➤ 11-20 studies	75€	150€	30€
<b>ZA CD Rom - customized</b> Data and documentation customized to user needs <u>and provided via FTP or e-mail</u>			
➤ Up to 5 studies	20€	30€	15€
➤ 6-10 studies	25€	50€	15€
➤ 11-20 studies	50€	100	20€
<b>ZA Download - online</b>	Free	Free	Free

### 2.1.6 Services offerts

Le **ZA** s'occupe de l'archivage de données primaires (données, questionnaires, variables) et de données sociales empiriques pour les préparer à des analyses secondaires et les rendre accessibles au public. Il est aussi expert dans les procédures techniques de la recherche sociale empirique et historique.

En plus de ses missions primaires, les activités et les services du ZA sont nombreux :

- Activités de recherche et développement dans le domaine de l'archivage, la conservation des données et leurs analyses (écoles d'été, publications, formations de jeunes chercheurs à l'archivage, ...);
- Envoi de supports documentaires aux utilisateurs;
- Echanges de données avec d'autres pays germanophones ou avec l'étranger via des organisations internationales ;
- Rédaction d'une publication bi-annuelle sur les tendances en recherche sociale ;
- Organisation de conférences et de conventions.

Des services pour les dépositaires de données comme le contrôle de leur donnée, l'archivage et le nettoyage de leurs données, mais aussi le dépôt des recherches sur le portail et l'accès gratuit au matériel disponible sur le portail sont également offerts par le ZA.

Enfin, le **site Internet** du ZA contient de nombreux liens vers des catalogues documentaires et des bases de données existantes en Allemagne et l'Europe (SOLIS, SOFIS, SOFO, ...<sup>10</sup>), et fait également des liens vers quelques logiciels d'analyse ou de classification de données<sup>11</sup>.

<sup>10</sup> <http://www.gesis.org/en/information/index.htm>

<sup>11</sup> <http://www.gesis.org/en/software/index.htm>

### 2.1.7 Interface

Pour terminer, voici quelques informations sur l'interface du ZA.

La page principale du portail (<http://www.gesis.org/en/za/index.htm>) présente des sujets très divers : des liens vers les catalogues de recherche, des données d'étude, des nouveautés... Le portail est accessible en anglais et en allemand, malgré le fait que toutes les ressources ne soient pas encore accessibles en anglais à ce jour.

Si le portail contient beaucoup d'information et de fonctionnalités (liens vers des catalogues, des logiciels d'analyse, possibilité de commander les documents en ligne, ...), il est assez difficile de s'y retrouver dans tout ce qu'il propose. Chaque catalogue possède en effet son propre moteur de recherche.

**Le catalogue de données générales**, le ZA Data Holding Catalogue, permet une recherche simple ou une recherche avancée.

Il est donc ainsi possible de **faire des recherches**

- Par champs prédéfinis dans le système : titre de la recherche, numéro d'étude, date, investigateur principal, collecteur de données, abstract, univers, mode de collecte des données, systèmes d'analyse, publications, notes, classification des sujets et disponibilité des données (recherche possible dans plusieurs champs pour la recherche avancée).
- Par langue de la description de l'étude : anglaise ou allemande
- Par classement ascendant ou descendant (pour les numéros de classification des études, les titres ou les dates de collecte des données).

A toute première vue, les références des données recherchées se présentent sous forme d'une liste sommaire (numéro de l'étude, titre, date de collecte des données). En cliquant sur chaque référence, il est possible d'accéder à une fiche plus détaillée contenant les **méta-données** plus détaillées comme le numéro de l'étude, le titre, la période de collecte des données, le commanditaire, le chercheur de données, un abstract de l'étude, l'univers de la recherche, la méthode utilisée, le mode de collecte de données (questionnaires, interviews, ...). Ces méta-données contiennent aussi des "données sur les données", c'est-à-dire le nombre de données (unités), de variables, les types de données et leur système d'exploitation. Enfin, la liste des publications associées à la recherche, des notes, la classification de l'étude (thème général dans lequel elle se trouve), des liens externes et la catégorie d'accès aux documents sont repris sur la fiche des méta-données.

Un inconvénient de ce portail est que les fiches de méta-données des recherches ne sont pas directement reliées à un formulaire de commande de documents. Au contraire, la partie téléchargement et commande des documents se situe dans une toute autre rubrique, la section "order and download", ce qui s'avère beaucoup moins pratique pour l'utilisateur.

### 2.1.8. Perspectives d'avenir

L'augmentation du nombre de visiteurs dans le portail et l'augmentation des données témoignent de l'utilité et de l'importance du service d'archivage allemand ZA ces dernières années. Renforcer la culture de partage de données chez les dépositaires, la collaboration avec d'autres centres d'archives, et la création des relais internationaux semblent des éléments essentiels assurant la pérennité de ce service d'archivage.

## 2.2. L'EXPERIENCE DE LA GRANDE-BRETAGNE: THE UK DATA ARCHIVE (UKDA)

### 2.2.1. Création et évolution<sup>12</sup>

La **naissance** de l'UKDA (<http://www.data-archive.ac.uk/>) remonte au premier octobre 1967, ce qui en fait le troisième centre d'archivage européen. Il fut créé à l'initiative du "Social and Economic Archive Committee", un comité destiné à favoriser l'échange d'informations sur les enquêtes en sciences sociales et des données générées par celles-ci. Les solutions proposées dans cette optique furent d'une part l'établissement d'un répertoire des données de recherche disponibles pour une réutilisation, et d'autre part la création d'un centre d'archivage. Celui-ci, financé par le Social Science Research Council (SSRC), fut établi au sein de l'Université d'Essex, qui fournissait le personnel qualifié et les locaux.

Il est intéressant de constater qu'à l'époque de sa création, le **but principal** du centre d'archivage était d'offrir aux chercheurs un accès aux données issues des enquêtes de marché ou d'autres sources commerciales, plutôt que des données issues de la recherche universitaire. Il fallait notamment éviter que ces données, par manque d'espace de stockage, ne soient revendues aux Etats-Unis et rendues ainsi indisponibles pour les chercheurs nationaux. Par la suite, l'UKDA étendit sa collecte à un nombre croissant de données issues de la recherche et dès les années 1970, il chercha à acquérir des ensembles de données d'origines gouvernementales.

### 2.2.2. Statut actuel, organisation et financement

L'UKDA n'est pas une entité juridique, c'est une partie de l'**Université d'Essex**. Il envisage aujourd'hui de devenir indépendant. Son statut actuel est le résultat d'une évolution historique et le centre n'a plus de sens de dépendre de l'Université, à l'heure actuelle. En être dépendant présente cependant l'avantage de pouvoir se reposer sur des ressources plus importantes à faire valoir en cas d'action en justice intentées contre le centre.

Le but actuel de l'UKDA est de rencontrer la demande en données. C'est un **"data broker"**. Il fait circuler des données, notamment vers l'étranger (en collaboration avec l'Université de Manchester), et offre un accès à des ensembles de données importants qui ne sont normalement disponibles que contre rémunération (données de la banque mondiale, du fond monétaire international,...). Cette dernière activité est une des justifications les plus importantes de son existence.

L'UKDA est divisé en **services multiples**: c'est une "umbrella organisation". L'ESDS, l'"Economic and Social Data Service", en constitue l'élément principal. En réalité, le projet SODA cherchera à étudier la faisabilité d'un équivalent belge de l'ESDS plutôt que de l'UKDA. L'UKDA réalise en outre un grand nombre d'activités pour différents organismes (l'Union Européenne,...).

L'UKDA emploie à l'heure actuelle environ l'équivalent de **50 temps plein**. Son personnel comprend environ 40% d'académiques et d'experts en sciences sociales, 20% d'informaticiens et 15% de personnel administratif et de gestion<sup>13</sup>.

---

<sup>12</sup> Information en provenance de *ACROSS THE DECADES - 40 years of data archiving*, brochure distribuée par l'UKDA à l'occasion de ses 40 ans, dont le contenu est disponible en ligne à l'adresse <http://www.data-archive.ac.uk/ukda40/>

<sup>13</sup> L'UKA est ainsi composée d'un staff technique (web-programmeurs, programmeurs d'enregistrement, systems manager) et de nombreux type d'autres personnes comme des assistants systèmes, des

L'ESDS, la section de l'UKDA véritablement consacrée à l'archivage pour les sciences sociales est structurée en **différent services** dont quatre services spécialisés dans le support pour les utilisateurs de certains types particuliers de données, un service principal de préservation et de dissémination, et un service de gestion faisant le lien entre les différentes parties. Le Directeur de l'ESDS reçoit le soutien du « Senior Management Team » rassemblant les dirigeants des cinq services (les services spécialisés et le service principal). Cette équipe se réunit régulièrement en personne ou par vidéoconférence.

L'ensemble de l'organisation rend des comptes aux organisations de financement quatre fois par an, et est également sous la supervision d'un « **Advisory Committee** »<sup>14</sup> qui réunit sur base biannuelle les représentants de toutes les parties intéressées dans le service : les organismes de financement, les utilisateurs, les déposants, les services de données liés, les services gouvernementaux de statistiques, et les autres professionnels de l'utilisation des données<sup>15</sup>.

L'ESDS est **financé** par le Social Science Research Council et travaille en collaboration avec le Joint Information Systems Committee (JISC)<sup>16</sup>, organismes publics de financement scientifique. Pour le reste de l'organisation, l'UKDA répond aux demandes de service de la part de ceux qui ont besoin de son aide pour conserver les données et trouve son financement dans ces sources diverses.

### **2.2.3. Type de données gérées**

#### **A. SELECTION**

Les données collectées par l'UKDA sont, de façon générale, des **données en “sciences sociales”**. Le terme n'est pas défini précisément car une telle définition semble irréalisable.

L'UKDA de façon générale conserve les sets de données qui lui sont confiées pour archivage. En ce qui concerne l'ESDS, un “**Acquisition Review Committee**” se charge de la validation du contenu. Ce comité comprend six personnes, dont deux s'occupent à temps plein des acquisitions. Celles-ci prennent en charge la sélection parmi les données spontanément offertes et la recherche active de données intéressantes. Les quatre autres membres du comité proviennent d'autres services et peuvent ainsi apporter un regard différent sur l'opportunité d'acquérir les données. L'Acquisition Review Committee se distingue de l'**Advisory Council**, lequel décide, deux fois par an, de l'orientation générale du centre.

La **sélection effectuée par les membres de l'Acquisition Review Committee** se base sur :

- les demandes des chercheurs, qui souhaitent obtenir tel ou tel set de données ; il y a alors une évaluation de l'importance de ces besoins ;
- les envois de données par les chercheurs. Ces envois sont acceptés s'ils en valent la peine, autrement dit s'il y a une réelle possibilité de réutilisation de leurs données. A l'heure actuelle, l'UKDA rejette deux tiers des données qui lui sont envoyées par des chercheurs financés par le JISC. C'est le cas notamment parce que de nombreuses recherches offrent toutes les données intéressantes à réutiliser dans la publication elle-

---

assistants web et des responsables de la digitalisation, des créateurs de méta-données, des assistants d'acquisition de données, ...

<sup>14</sup> La liste des membres de l'Advisory Committee se trouve à l'adresse

<http://www.esds.ac.uk/about/advisory.asp>

<sup>15</sup> *Economic and Social Data Service - Strategic Plan 2004-2009*, p.17

<sup>16</sup> The mission of the Joint Information Systems Committee (JISC) is to provide world-class leadership in the innovative use of ICT to support education and research. <http://www.jisc.ac.uk/>

même, les autres données ne présentant pas d'intérêt. Cependant, pour l'avenir, l'UKDA est occupé à mettre sur pieds un **système de « self deposit »** qui leur permettra d'accepter toutes les offres.

## B. CARACTERISTIQUES

Les données conservées sont de type :

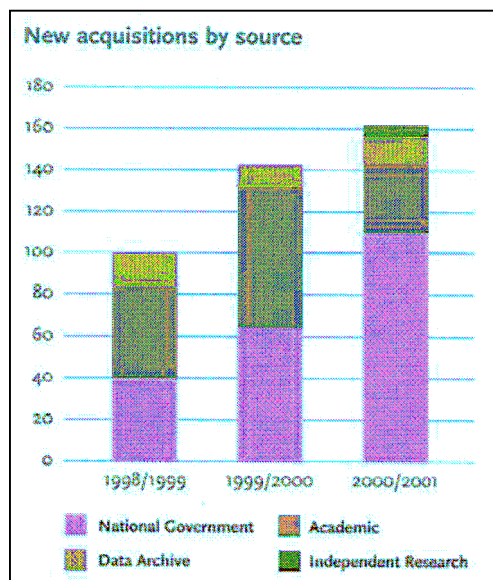
- **quantitatives** : micro données (réponses individuelles aux enquêtes codées numériquement), macro données (données agrégées ex : niveau économique d'un pays). Format des données en SPSS, Stata et Tab
- **qualitatives** : interviews en profondeur, journaux, notes anthropologiques et réponses complètes aux questions d'enquêtes. Formats sous forme de fichiers Excel, Word et RTF
- **multimédia** : un petit nombre de données comprennent des fichiers image, tels que des photographies et des clips audio
- **matériel non digital** : du matériel papier (photographies, rapports, questionnaires et transcriptions), du matériel audio analogue ou des enregistrements audio-visuels.<sup>17</sup>

L'UKDA ne conserve et ne fournit que des **données anonymisées**, sauf dans le cas où les répondants ont donné leur permission ou lorsque les données font partie du domaine public, ceci pour éviter de violer la vie privée des personnes interrogées et la confiance des déposants. Soit les données sont déjà anonymes à la collecte, soit elles sont anonymisées par le chercheur ou par le centre.

### 2.2.4. Collecte des données

#### A. SOURCE DES DONNEES

L'UKDA reçoit chaque année des données en provenance de différentes sources, y-compris :



- les agences officielles, principalement le gouvernement du Royaume-Uni (environ 60% des acquisitions)
- les agences statistiques internationales
- les chercheurs individuels à qui ont été octroyés des bourses de recherche (environ 30% des acquisitions)
- les agences d'études de marché
- les sources historiques
- d'autres centres d'archivage étrangers<sup>18</sup>.

<sup>17</sup> Source : <http://www.esds.ac.uk/findingdata/data.asp>

<sup>18</sup> Source : <http://www.esds.ac.uk/findingdata/data.asp> et Economic and Social Data Service – Annual report 2000-2001, p. 4, disponible à l'adresse <http://www.data-archive.ac.uk/news/publications/AnnRep00.pdf>

## **B. METHODES DE COLLECTE**

### ***Données publiques***

L'UKDA a conclu un accord (concordat) avec l'office national des statistiques, qui est responsable de coordonner les différents départements d'archivage. **L'Etat** est son fournisseur principal de données : environ 60% des ensembles de données proviennent d'agences officielles.

### ***Données universitaires***

#### **Comment l'UKDA parvient-il à convaincre les chercheurs d'y déposer leurs données ?**

La réponse est simple : en les y obligeant dans leurs contrats de financement. Par la suite, on peut les convaincre en faisant valoir la visibilité. C'est ainsi que tous les contrats de financement de l'Economic and Social Research Council et du JISC contiennent une clause obligeant les chercheurs à offrir leurs données à l'UKDA.

Dès qu'un nouveau projet est financé, l'UKDA en reçoit la description et le catégorise en fonction de l'intérêt des données qui en sortiront. Le centre écrit ensuite au chercheur pour l'informer de cette catégorisation et pour lui rappeler ses obligations générales en matière de confidentialité, de l'importance d'une bonne documentation, etc. Cette démarche est très intéressante car les chercheurs sont rarement au courant de leurs obligations liées à la confidentialité, et ceci permet de leur éviter des problèmes par la suite, notamment des problèmes d'archivage. Une autre justification de l'utilité du centre est que celui-ci fournit un support qui aide à la création des données suivant les meilleurs standards.

### ***Données privées***

En ce qui concerne les **données privées**, l'UKDA n'a pas d'accords formels avec des entreprises. Le centre récolte de moins en moins de données issues d'entreprises (les enquêtes d'opinion, etc) car elles sont souvent difficiles à utiliser, méthodologiquement mal réalisées, et rencontrent peu de demandes de la part des scientifiques. Cependant, ce n'est pas le cas de toutes les données privées, et l'un des buts de l'UKDA est d'accéder aux données privées de marketing (par exemple, les statistiques d'achat établies par les supermarchés), aux données historiques, etc.

#### **Pour récupérer les données privées intéressantes, l'UKDA emploie deux techniques :**

1. l'achat de ces données ;
2. persuader les entreprises qui ont financé l'enquête que partager les données est dans l'intérêt de la recherche et dans leur propre intérêt :
  - des nouvelles idées pourront naître dans le milieu scientifique à partir de ces données, et les entreprises en bénéficieront ;
  - les données seront conservées de façon sûre ;
  - lors de la réutilisation de ces données, une publicité sera faite à l'entreprise ; c'est de la publicité à bon marché.

## **C. LICENCE AGREEMENT**

Au moment du dépôt, la signature d'un contrat est imposée par l'UKDA au déposant : le « **licence agreement** ».

Le contrat s'adresse au « depositor », qui est le titulaire du « copyright » sur la base de données (l'auteur), le co-titulaire, ou un tiers autorisé par le(s) titulaire(s).

**Droits accordés au depositor :**

- droit d'utiliser ou de publier les données ailleurs; le licence agreement ne transmet pas de droits d'auteur ;
- garantie que le ré-utilisateur sera informé que les données ne peuvent être utilisées sans la permission de l'UKDA ;
- garantie que le ré-utilisateur sera informé de son obligation, dans toute publication sur la base des données réutilisées, d'insérer une clause de reconnaissance (acknowledgment) nommant l'auteur des données et celui qui les a financées ;
- conservation des données, sans obligation de résultat.

**Obligations du depositor :**

Le « depositor » n'apporte aucune garantie concernant les données en termes de portée (comprehensiveness), de précision, de fiabilité, ou autre.

Par contre, il garantit :

- qu'il est autorisé à accorder les droits spécifiés dans le contrat et que ces droits ne violeront pas les droits de tiers ;
- que la collection de données respecte les lois en vigueur ;
- qu'aucune obligation légale ou contractuelle n'est applicable au depositor qui l'empêcherait de s'engager ou de remplir ses obligations prévues par le contrat;
- qu'il informera l'université de tout changement de propriété intellectuelle des données;
- qu'il informera l'université de toute question de confidentialité, de protection de la vie privée ou de protection des données relative à la collection de donnée.

De plus, il s'engage à indemniser l'université en cas de recours suite à une négligence, une omission ou une violation du contrat de sa part.

**Obligation de la plateforme :**

En contrepartie, la plateforme doit s'assurer que l'accès aux données se fasse dans le respect des conditions décrites dans le licence agreement. Ces conditions prévoient les types d'accès autorisés (*cf. infra*).

## **2.2.5. Accès aux données**

L'UKDA regroupe de nombreuses données d'origine gouvernementale qui sont extrêmement intéressantes d'un point de vue scientifique, mais pour les obtenir, il faut pouvoir garantir que la confidentialité sera respectée et offrir un niveau de sécurité élevé. Dans ce but, l'UKDA "coache" les données via:

- le "licence agreement" avec le déposant (*cf. supra*);
- l'enregistrement des utilisateurs, qui doivent adhérer à un « set of terms and conditions » ;
- un suivi des utilisations par les différents utilisateurs (via les données d'enregistrement) ;
- pour les données présentant un risque plus élevé, une "special licence".

### **A. L'ENREGISTREMENT**

Le **système d'enregistrement** choisi est doublé d'un contrôle effectif des informations fournies:

- l'authentification se fait via le système "Athens", où sont enregistrés tous les académiques de Grande-Bretagne;

- les autres utilisateurs doivent remplir un formulaire et demander la création d'un compte dans Athens;
- ce système permet d'avoir des statistiques précises concernant l'utilisation du centre;
- il permet aussi de renforcer la confiance des déposants (publics ou non), parce qu'il est possible de retrouver les responsabilités en cas de mauvaise utilisation. Il est impossible de garantir à 100% que l'utilisation des données sera conforme aux règles, mais grâce à ce système, l'UKDA peut prouver sa diligence dans la sécurisation des données, et peut entamer des actions en justice envers les coupables de mauvaises utilisations.

## B. CONTRAT DE REUTILISATION

Les conditions d'accès aux données sont spécifiées par les dépositaires de celles-ci. Le dépositaire a le choix entre des conditions standards d'accès aux données ("End User Licence") et des conditions spéciales d'accès aux données ("Special licence").

Au moment de l'enregistrement, l'utilisateur doit adhérer aux termes du « **end user licence** ».

Aux termes du « end user licence », le réutilisateur a les **obligations suivantes** :

- Préserver la confidentialité des données : ne pas chercher à identifier les données.
- Citer la source, pour que d'autres chercheurs puissent retrouver facilement les données, et reconnaissance (acknowledgment) contenant les informations de copyright.
- Ne pas faire de transfert des données : interdiction de transmettre les données reçues à quiconque n'est pas enregistré sur UKDA et/ou n'a pas adhéré aux conditions d'accès pour le set de données.
- Préserver la sécurité des données et des mots de passe fournis.
- Mettre les données dérivées à disposition de l'UKDA.

Certaines données sont directement disponibles sous le contrat « end user licence », tandis que d'autres nécessitent une « special licence ».

Les premières données sont totalement anonymes et certaines variables peuvent être supprimées pour limiter le risque d'atteintes à la confidentialité (exemple : les variables géographiques). Le « end user licence » peut être signé par tout utilisateur enregistré qui s'engage à n'utiliser les données que pour des buts d'enseignement, d'apprentissage ou de recherche sans but lucratif.

Les secondes sont également anonymes mais plus détaillées, et présentent donc plus de risques d'atteintes à la confidentialité. Le « special licence » ajoute des conditions à leur accès :

- l'utilisateur doit démontrer qu'il a besoin de ces données, à l'exclusion d'aucune autre;
- il doit prouver qu'il les utilise dans un environnement privé ;
- il doit soumettre toutes les publications à l'avance.

## C. TYPES D'ACCES

Les données sont disponibles pour les buts déterminés par le déposant :

- soit les données ne sont disponibles que pour des buts non-commerciaux et de recherche;
- soit elles sont disponibles sans restrictions (but commerciaux et non-commerciaux) ;

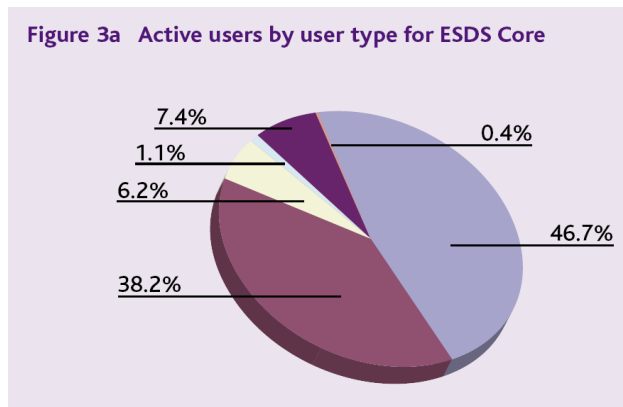


- soit elles sont disponibles pour des buts non-commerciaux et pour des buts commerciaux moyennant le paiement de royalties.

Lorsque les données sont disponibles, elles le sont simultanément pour la réutilisation, la copie, l'accès total ou partiel.

#### D. TYPES DE REUTILISATEURS<sup>19</sup>

Figure 3a Active users by user type for ESDS Core



Les données sont offertes dans un but principalement scientifique et académique ; par conséquent, la majorité des utilisateurs font partie du monde de l'enseignement et de la recherche. L'UKDA ne propose que très peu d'utilisation commerciale, qui comporte plus de difficultés que d'avantages ; son rôle n'est pas de tirer profit des données.

Legend for Figure 3 charts

academic postgraduate undergraduate other student public sector commercial

#### E. MESURES DE SECURITE

Le système de sécurité utilisé par l'UKDA est nommé **"Shibboleth"** ; c'est un « open source authentication service ». Dans quelques temps, il remplacera Athens. Pour le moment, les deux coexistent. Les raisons de ce remplacement sont les suivantes :

- Shibboleth est plus complexe qu'Athens, qui offre un nombre limité de paramètres d'accès. Avec Shibboleth, la réponse à l'accès n'est plus oui ou non mais des accès modulables sont possibles.
- Il est également réellement international (contrairement à Athens, qui est purement national) car il peut communiquer avec d'autres systèmes. Le but est de rendre possible un enregistrement unique pour tous les pays, sous la forme d'un passeport: l'enregistrement dans le pays A sera reconnu comme moyen d'accès dans les autres pays.

#### 2.2.6. Services offerts

L'UKDA a pour but principal la collecte, la conservation, la vérification, la préparation pour réutilisation et la dissémination des données qui lui sont confiées. Il offre cependant un certain nombre de services supplémentaires :

- La création d'un répertoire des données disponibles ;
- La tenue d'une « bibliothèque virtuelle » offrant un accès direct en ligne à un nombre croissant de documents ;
- Des activités de recherche et de développement (recherche de nouvelles méthodes de conservation et de publication des données, de standards de métadonnées, de softwares,...)<sup>20</sup> ;

<sup>19</sup> Source des données : *Economic and Social Data Service – Annual report 2003-2004*, p.20, disponible à l'adresse <http://www.data-archive.ac.uk/news/publications/annrep2004.pdf>

- Des activités d'enseignement, en offrant des formations sur les méthodes de recherche, de classement ou d'analyse des données ;
- Des activités de support et de conseils pour le traitement des données en vue d'une réutilisation (help desk et conseils en ligne, outils d'analyses téléchargeables, contacts avec les chercheurs,...)<sup>21</sup>.

### **2.2.7. Interface**

Sur le site Internet <http://www.data-archive.ac.uk/>, les données humaines et sociales gérées par l'UKDA sont classées en **quatre catalogues** :

- l'ESDS Gouvernement, comprenant les enquêtes gouvernementales d'échelle nationale ;
- l'ESDS International, reprenant des données internationales comme celles issues des Eurostats, Eurobaromètres, World Bank Data, World Value Survey,...
- l'ESDS Longitudinal, le service d'études longitudinales
- l'ESDS Qualidata, un service particulier pour certains ensembles de données qualitatives.

Les données sont présentées en ligne sous forme de fiches documentaires respectant les conventions internationales d'archivage décrivant les métadonnées. L'utilisateur peut accéder aux descriptions des ressources soit en passant par le catalogue général (data catalogue), soit en explorant les « major studies ». Une recherche est possible par titre, mot clé, date de création, lieu géographique, numéro de l'étude, publication et variable.

**L'accès** aux fiches documentaires et à leur documentation (guide de l'utilisateur, informations sur l'étude, déclarations de variables, ...) est totalement libre et ne demande pas d'enregistrement. Par contre, l'accès aux données brutes requiert un enregistrement et une authentification de l'utilisateur. De plus, à chaque demande de données brutes, il est demandé de préciser les buts de l'utilisation.

Le site web fournit également des informations sur l'actualité du service, des conseils d'utilisation, des liens vers d'autres catalogues et banques de données, des annonces de formations possibles, des outils gratuits téléchargeables, ainsi qu'un accès au système Nesstar offrant la possibilité de voir directement les données en ligne et de les analyser via des croisements.

### **2.2.8. Perspectives d'avenir**

Les buts pour le futur de l'UKDA sont de continuer à étendre le **nombre des données** collectées et la **taille du catalogue**, ainsi que de faciliter **l'accès en ligne**. De plus, dans un environnement de données globalisé, le centre cherche à créer de nouveaux outils de transmission des données et à favoriser la collaboration internationale, dans le but de permettre à tout chercheur d'obtenir les données dont il a besoin, où qu'elles se trouvent.<sup>22</sup>

---

<sup>20</sup> Plus d'informations dans *ACROSS THE DECADES - 40 years of data archiving*, brochure distribuée par l'UKDA à l'occasion de ses 40 ans, dont le contenu est disponible en ligne à l'adresse <http://www.data-archive.ac.uk/ukda40/> (p.30)

<sup>21</sup> *Idem*, p.22

<sup>22</sup> *Idem*, p.36

## Chapitre 3. Le contexte international

---

L'archivage et l'échange de données scientifiques et des sciences sociales en particulier a été l'objet d'une attention particulière au sein de la communauté internationale ces dernières années. Dans ce chapitre sont retracées les déclarations politiques les plus importantes en la matière, ainsi que les initiatives concrètes au niveau international.

### 3.1. LA POSITION DE LA COMMUNAUTÉ INTERNATIONALE

#### 3.1.1. La déclaration de Berlin

La *Déclaration de Berlin sur le Libre Accès à la Connaissance en Sciences exactes, Sciences de la vie, Sciences humaines et sociales*<sup>23</sup> a été signée le **22 octobre 2003** par dix-neuf hauts représentants du monde scientifique international. Depuis lors, le nombre des signataires s'est accru pour atteindre **242 organisations** actives dans l'enseignement, la recherche, le financement scientifique, associations privées ou autorités publiques. A l'heure actuelle, dix-neuf organisations, autorités et universités belges se sont engagées à respecter les termes de la déclaration<sup>24</sup>.

La déclaration de Berlin, héritière de la Déclaration de Budapest pour l'accès ouvert, de la charte ECHO et de la Déclaration de Bethesda pour l'édition en **libre accès**, cherche à « promouvoir un Internet qui soit un instrument fonctionnel au service d'une base de connaissance globale et de la pensée humaine ». Elle s'adresse aux responsables politiques en charge de la science, aux institutions de la recherche, aux agences de financement, aux bibliothèques, aux archives et aux musées, et formule les mesures à envisager pour contribuer au libre accès de la connaissance.

Selon les termes de la déclaration, « les contributions au libre accès se composent de résultats originaux de recherches scientifiques, de données brutes et métadonnées, de documents sources, de représentations numériques de documents picturaux et graphiques, de documents scientifiques multimédias ». Ces contributions doivent satisfaire à deux conditions :

1. « Leurs auteurs et les propriétaires des droits afférents concèdent à tous les utilisateurs un droit gratuit, irrévocable et mondial d'accéder à l'œuvre en question, ainsi qu'une licence les autorisant à la copier, l'utiliser, la distribuer, la transmettre et la montrer en public, et de réaliser et de diffuser des œuvres dérivées, sur quelque support numérique que ce soit et dans quelque but responsable que ce soit, sous réserve de mentionner comme il se doit son auteur (...) ».
2. Une version complète de cette œuvre, ainsi que de tous ses documents annexes (...) est déposée (et, de fait, publiée) sous un format électronique approprié auprès d'au moins une archive en ligne, utilisant les normes techniques appropriées (comme les définitions des Archives Ouverts [Open Archives]), archive gérée et entretenue par une institution académique, une société savante, une administration publique, ou un organisme établi ayant pour but d'assurer le libre accès, la distribution non restrictive, l'interopérabilité et l'archivage à long terme. »

---

<sup>23</sup> Disponible à l'adresse [http://oa.mpg.de/openaccess-berlin/BerlinDeclaration\\_wsisis\\_fr.pdf](http://oa.mpg.de/openaccess-berlin/BerlinDeclaration_wsisis_fr.pdf)

<sup>24</sup> Liste des signataires disponible à l'adresse <http://oa.mpg.de/openaccess-berlin/signatories.html>

La création d'une archive ouverte en sciences sociales permettrait aux signataires belges de rencontrer idéalement leurs engagements dans ce domaine. Cela implique cependant, comme l'indique la déclaration, de « trouver des solutions aptes à soutenir le développement es cadres juridique et financier actuels en vue de faciliter un accès et un usage optimaux ».

### **3.1.2. La position de la Commission européenne**

La Commission européenne a également souligné récemment l'importance d'un large accès à l'information scientifique. Dans sa "*Communication au Parlement européen, au Conseil et au Comité économique et social européen sur l'information scientifique à l'ère numérique : accès, diffusion et préservation*"<sup>25</sup>, la Commission insiste sur l'importance du système de **publication scientifique** et son **impact sur l'excellence scientifique**. Elle déclare « lancer un processus politique sur (a) l'accès et la diffusion de l'information scientifique et (b) les stratégies de préservation de l'information scientifique dans toute l'Union Européenne » et invite le Parlement européen et le Conseil à débattre de ces problématiques, dans la lignée de la Déclaration de Berlin.

## **3.2. LES RÉSEAUX D'ÉCHANGE DE DONNÉES**

### **3.2.1. Council of European Social Science Data Archive (CESSDA)**

#### **A. BUT**

Le CESSDA (<http://www.cessda.org>) est une organisation regroupant les centres d'archivage nationaux de données en sciences sociales au niveau européen. Le **but** de CESSDA est de construire une **infrastructure légale et technique** pour la **création d'un réseau d'échange de données**. Il peut être également envisagé sous une forme décentralisée comme un ensemble de services séparés partageant un environnement commun.

Ses objectifs concrets sont définis à l'article 1 de sa charte de fonctionnement :

- « 1. Promouvoir l'acquisition, l'archivage et la distribution de données à travers l'Europe.
2. Promouvoir des projets et des procédures pour accroître l'échange de données et de technologies entre les organisations de données.
3. Stimuler le développement et l'utilisation de ces procédures à travers l'Europe.
4. Encourager les nouvelles organisations de données à suivre ces objectifs.
5. Promouvoir l'intégration d'une base de données européenne.
6. S'associer et coopérer avec d'autres organisations internationales partageant ces objectifs. »<sup>26</sup>

#### **B. FONCTIONNEMENT**

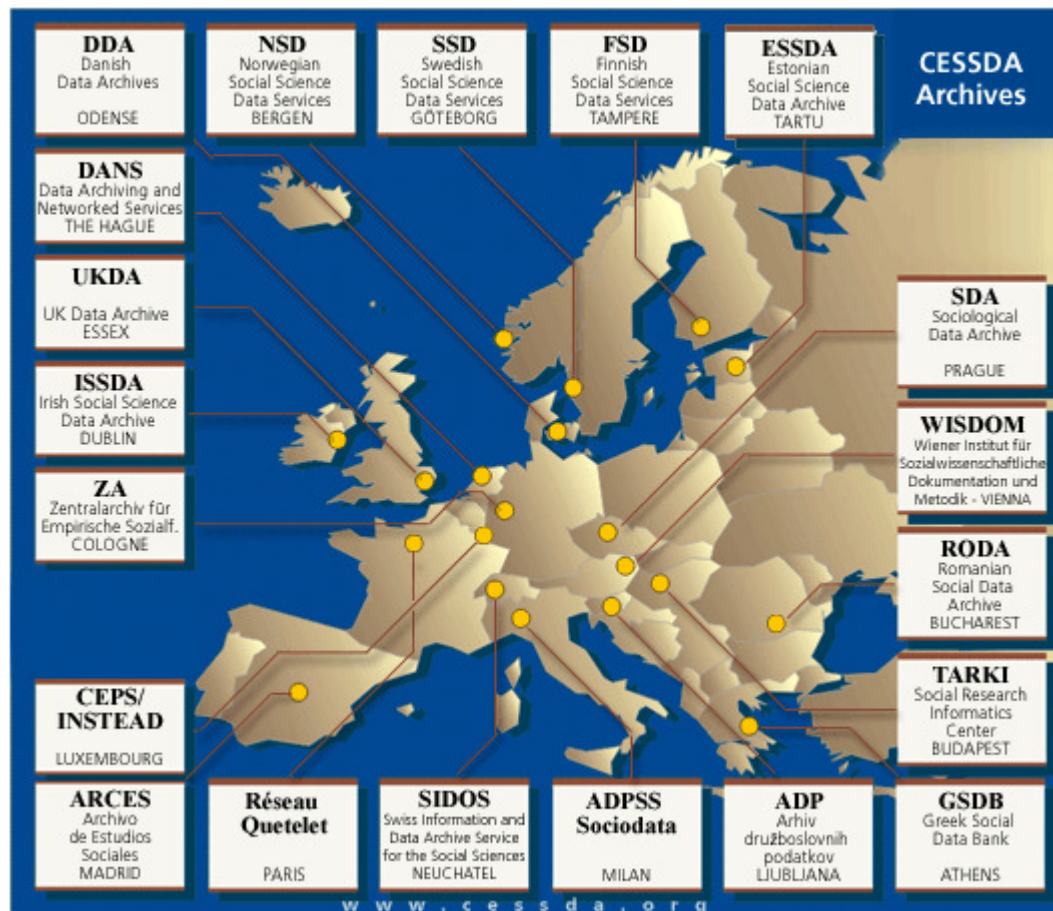
Le CESSDA n'est pas une organisation en tant que telle ; c'est un réseau qui ne possède pas d'entité juridique. Son fonctionnement est réglé par ses « Articles of governance ». Il compte actuellement une vingtaine de membres, répertoriés sur son site web<sup>27</sup>.

<sup>25</sup> COM(2007) 56 final, 14 février 2007, disponible à l'adresse [http://ec.europa.eu/research/science-society/document\\_library/pdf\\_06/communication-022007\\_fr.pdf](http://ec.europa.eu/research/science-society/document_library/pdf_06/communication-022007_fr.pdf)

<sup>26</sup> Traduction libre de l'article 1 - CESSDA, *Articles of governance*, disponible à l'adresse <http://extweb3.nsd.uib.no/opencms7final/export/sites/default/cessda/pdf/cessdaconstitution20040402.pdf>

<sup>27</sup> <http://www.cessda.org>

## CESSDA Member Archives



Le CESSDA est composé d'une **Assemblée Générale** (avec un représentant par membre) qui se réunit chaque année, et d'un **Comité Exécutif** (de trois à cinq représentants) qui prend en charge la gestion journalière. Monsieur Kevin Schürer, actuel directeur de l'UKDA, en est actuellement le président.

Le **fonctionnement** du CESSDA est financé par une contribution de ses membres. Le réseau postule en outre pour l'obtention de financement de la part de l'Union Européenne en vue du développement de nouveaux outils d'échange de données.

### C. ADMISSION

Jusqu'à présent, aucun critère d'admission officiel pour les centres d'archivage membres du CESSDA n'ont été reconnus. Cependant, un « membership criteria committee » a été récemment mis sur pieds et est en train de rédiger un document à ce sujet. La politique actuelle du CESSDA semble tendre vers une admission plus sélective, avec plusieurs niveaux de participation :

- « full member » : accessible aux centres qui stockent les données dans un environnement interopérable, et dont les données sont répertoriées dans le catalogue de CESSDA. Tous les membres effectifs devront partager le même catalogue, un contrat avec l'utilisateur commun et une méthode d'authentification commune.

- « associate member » ou « affiliated member » : un type de participation moins exigeant, réservé aux centres d'archivages qui auraient des accords avec le CESSDA sans répondre aux conditions d'un « full member ».

Le CESSDA n'admet que des centres d'archivage à portée nationale. Cela ne signifie pas qu'il ne pourra y avoir qu'un **seul membre par pays** : le réseau peut comprendre plusieurs membres dans le cas où l'archivage de données est décentralisé, à la condition que chaque membre soit le seul détenteur au niveau national du type de données dans lequel il est spécialisé.

#### **D. ACCORD D'ACCÈS TRANSFRONTALIER AUX DONNÉES**

Pour assurer l'échange de données entre les membres, un « **Trans-border Data Access Agreement** » a été établi et imposé à tous les membres<sup>28</sup>. Il prévoit un échange de données entre membres à la demande de chercheurs étrangers sans coûts additionnels et avec un accès équivalent à celui des chercheurs nationaux. Chaque membre choisit s'il veut satisfaire à cette obligation en étant directement en contact avec le chercheur ou via le centre du pays où se trouve le chercheur. Dans tous les cas, les membres du CESSDA ne doivent pas conserver les données fournies par un autre membre, sauf avec leur accord si les données sont pertinentes pour leur propre pays.

#### **E. INTERFACE**

Dans son rôle d'échange des données conservées par les centres nationaux, le CESSDA offre l'accès à un portail-carrefour permettant d'effectuer une recherche dans les catalogues de ses membres. Le site internet du CESSDA propose un thesaurus commun à l'ensemble de ceux-ci. Le thesaurus utilisé est l'ELSST, pour European Language Social Science Thesaurus. Il a fait l'objet de traductions dans la plupart des langues de l'Union européenne.

La recherche dans le portail CESSDA se fait soit en recherche libre, soit à partir du thesaurus ELSST, soit directement dans le catalogue d'un centre membre. Le moteur de recherche est très performant.

#### **3.2.2. International Federation of Data Organizations for the Social Science (IFDO)**

L'IFDO (<http://www.ifdo.org>) est l'équivalent du CESSDA au niveau mondial : un réseau d'échange de données rassemblant des centres d'archivages nationaux en sciences sociales. L'IFDO compte à l'heure actuelle 35 membres répartis dans le monde entier<sup>29</sup>.

L'IFDO travaille en étroite collaboration avec le CESSDA et présente une structure similaire. L'étude des différents projets se fait via des comités spécialisés, et l'ensemble de la structure est gérée par un secrétariat général. Le président actuel de l'IFDO est Ekkehard Mochmann, directeur du ZA.

L'IFDO est cependant moins ambitieux que le CESSDA : il ne cherche pas à unifier les portails mais à atteindre leur collaboration. De ce fait, il n'offre pas de catalogue en ligne des données

---

<sup>28</sup> Disponible en ligne à l'adresse

[http://extweb3.nsd.uib.no/opencms7final/export/sites/default/cessda/pdf/tbaa\\_v23.pdf](http://extweb3.nsd.uib.no/opencms7final/export/sites/default/cessda/pdf/tbaa_v23.pdf)

<sup>29</sup> Liste des membres disponible à l'adresse <http://www.ifdo.org/network/index.html>

disponibles dans le réseau, mais propose une série de liens vers des bibliothèques virtuelles tenues par ses membres<sup>30</sup>. Le site offre en outre une série de documents contenant des informations pratiques sur l'archivage et la gestion de données en sciences sociales.

### **3.2.3. East European Data Archive Network (EDAN)**

En-dehors du CESSDA et de l'IFDO et parmi les autres réseaux internationaux, il existe un autre réseau européen organisé dans le domaine de l'échange de données en sciences sociales : le « **East European Data Archive Network** » (EDAN)<sup>31</sup>. Créé en 2002, il a pour but d'agir comme un réseau informel d'entraide entre les centres d'archivage d'Europe de l'Est qui sont encore au début de leur développement. Sous le patronage du Gesis allemand, il offre un cadre institutionnel aux nouvelles archives pour les aider à atteindre le niveau de développement de leurs aînées occidentales.

---

<sup>30</sup> <http://www.ifdo.org>

<sup>31</sup> [http://www.gesis.org/en/cooperation/data\\_service/eastern\\_europe/](http://www.gesis.org/en/cooperation/data_service/eastern_europe/)

## **2EME PARTIE : BESOINS ET ATTENTES DES ACTEURS AUTOUR DE L'ARCHIVAGE**

### **Introduction : Méthodologie**

Pour pouvoir proposer un projet de service d'archivage bien adapté au contexte particulier de la Belgique, il était essentiel d'interroger les promoteurs et utilisateurs potentiels du futur service d'archives. Outre les avis réguliers que nous ont transmis les membres du Comité d'Accompagnement de notre projet (composé de professeurs d'universités, de représentants de la recherche scientifique en Flandre et en Communauté française, de représentants du monde des archives ...), nous avons organisé deux groupes de discussion destinés à faire connaître notre projet et à mieux connaître les motivations et les attentes des personnes vis-à-vis du futur service d'archivage.

Deux focus-groups d'une demi-journée chacun ont donc eu lieu :

- Le 27 juin 2007 avec des représentants d'institutions administrant des programmes de recherche ou actives dans l'archivage (SPF Politique Scientifique Fédérale, représentants des administrations de la recherche scientifique, et des AGR : archives générales du royaume).
- Le 20 juillet 2007 avec des représentants du monde académique et scientifique (Chercheurs et Professeurs d'Université du domaine des sciences sociales).

Une fois de plus, toutes ces personnes ayant participé au groupe de discussion ne sont pas représentatives du monde des utilisateurs de l'archivage. Leurs réactions permettent cependant de se faire une idée de leurs réalités et de leurs besoins. Une enquête chiffrée à l'échelle de la Belgique pourrait donner une meilleure représentativité à cette étude.

## **Chapitre 1. Groupe de discussion avec les acteurs institutionnels**

### **1.1. DÉROULEMENT DU GROUPE DE DISCUSSION**

Ce groupe de recherche rassemblait des représentants de certaines administrations fédérales et fédérées en lien avec la recherche, l'archivage ou provenant d'autorités manipulant de grandes quantités de données réexploitables. Le but n'était pas de rassembler des avis officiels mais de soulever les souhaits et sensibilités dans un contexte informel. Le programme de ce groupe de discussion se trouve en **annexe 4** de ce rapport.

Les débats ont été divisés en trois phases :

1. Une discussion libre sur les motivations et les apports de la création d'un service d'archivage de données en sciences sociales, du point de vue des acteurs institutionnels ;



2. Une présentation de différentes combinaisons de fonctionnalités possibles et une réflexion commune sur leurs intérêts respectifs ;
3. L'évocation en commun de pistes de réflexions concernant la mise en place concrète et les implications institutionnelles dans le service.

Ces deux dernières phases visaient à mettre en lumière différents scénarios possibles qui seront évoqués dans la troisième partie de ce rapport.

## 1.2. MOTIVATIONS ET APPORTS DE LA CREATION DU SERVICE D'ARCHIVAGE

Les personnes interrogées ont largement reconnu **deux types de motivations** pour la création d'un centre d'archivage en sciences sociales en Belgique : des motivations scientifiques et des motivations politiques ou administratives.

Au niveau des motivations scientifiques, les discussions ont retenu l'intérêt de **l'échange et de la réutilisation des données**. Les participants ont fait valoir qu'un système permettant l'archivage et l'échange des données faciliterait le travail de nombreuses parties intéressées, et éviterait la perte inutile d'informations intéressantes pour des recherches ultérieures. Ce système aurait aussi comme effet d'offrir un **contrôle de la méthodologie et des procédures** en offrant un regard extérieur sur la qualité des recherches et des données récoltées ; il aurait ainsi un effet très positif sur le niveau de la recherche scientifique. D'après les personnes interrogées, le personnel scientifique s'inquiète de la **conservation à long terme des résultats d'enquêtes**, conservation qui ne peut être garantie qu'au prix d'un travail important ; un système d'archivage répondrait à des attentes à ce niveau-là également, ce qui profitera notamment aux administrations confrontées au problème de la conservation de leurs données. Enfin, la possibilité de donner aux chercheurs un **accès plus facile à la publication** via un archivage public a également été invoquée.

En ce qui concerne les **motivations politiques et administratives**, les personnes présentes ont souligné l'intérêt économique d'une **double valorisation** des recherches financées par des fonds publics ou privés. Un système d'archivage aurait un potentiel d'économie des coûts, via la réutilisation des données qui n'ont plus à faire l'objet de recherche. Il permettrait aussi de revaloriser toutes les étapes de la recherche et non pas uniquement ses conclusions, ce qui augmenterait le rendement du financement et pourrait également renforcer le principe démocratique. La mise à disposition des données intermédiaires et des métadonnées faciliterait également un **contrôle de la qualité** de la recherche, ce qui profiterait aux organismes de financement. Le souhait de récupérer via un centre d'archivage les **résultats de recherches privées** (comme les enquêtes Shell) pour les chercheurs privés ou pour les autorités publiques a aussi été évoqué. Enfin, l'intérêt pour la **visibilité de la recherche** a été souligné comme un atout majeur ; d'après les participants, la Belgique serait peu visible sur le plan scientifique parce qu'elle manque d'interlocuteur unique pour la valorisation de son potentiel scientifique, rôle qu'un centre d'archivage pourrait jouer partiellement.

Cependant, les participants ont souligné l'importance d'une **évaluation des apports effectifs de cet archivage**. L'investissement des autorités sera à la mesure du profit d'une telle initiative pour la communauté des chercheurs et pour les institutions elles-mêmes. Si les données récupérées ne sont pas directement intéressantes pour l'administration, elles peuvent l'être de façon indirecte, via l'intérêt des chercheurs ; mais cet intérêt doit être attesté pour garantir un retour sur l'investissement consenti.

## Chapitre 2. Groupe de discussion avec les acteurs scientifiques et académiques

---

### 2.1. DEROULEMENT DU GROUPE DE DISCUSSION

Ce groupe de discussion s'est déroulé en quatre phases axées sur :

- 1) Les motivations des participants pour la création d'un service d'archivage et les principaux apports de ce type de service.
- 2) La présentation et les fonctionnalités du service d'archivage.
- 3) Les conditions du dépôt de données dans ce portail d'archive (type de données à déposer, conditions posées pour les mettre à disposition dans un portail d'archive et obstacles possibles à cette mise à disposition).
- 4) Les conditions d'accès et de réutilisation des données du portail d'archive (type de données pouvant être réutilisées, utilisateurs potentiels, procédures d'accès aux données, obstacles possibles à cette mise à disposition).

A partir de la seconde phase, nous avons proposé plusieurs scénarios non limitatifs aux participants. Ceux-ci pouvaient choisir les scénarios qu'ils préféraient, mais également en proposer de nouveaux. Le programme de ce groupe de discussion se trouve en **annexe 5** de ce rapport. Avant de présenter les scénarios choisis par les participants (voir partie 3 du présent rapport), nous allons nous attarder sur leurs motivations en faveur du service d'archives.

### 2.2. MOTIVATIONS ET APPORTS DE LA CREATION DU SERVICE D'ARCHIVAGE

De manière générale, la proposition de création d'un service d'archivage pour les sciences sociales en Belgique a été très bien accueillie par les différents représentants de la recherche que nous avons rencontrés. En effet, ces différents représentants y ont vu plusieurs opportunités. La première serait de favoriser un **libre accès multidisciplinaire** à la connaissance et à la science au niveau national et international, comme le souligne actuellement la déclaration de Berlin<sup>1</sup>. Une autre opportunité largement citée serait la centralisation des données au sein d'un **guichet unique**. Il résulterait de ces deux opportunités plusieurs bénéfices importants pour le monde académique et scientifique : une **facilité** pour obtenir des données et les traiter (comparaisons, analyses, ...), un **décloisonnement** du secteur de la recherche, un **gain de temps**, moins de gaspillage de recherches réalisées, et plus largement une **meilleure efficacité** et une **meilleure rentabilité** de la recherche. Enfin, les participants ont souligné qu'un tel portail permettrait aussi d'améliorer la **visibilité de la recherche en sciences sociales** menées en Belgique sur la scène européenne et internationale.

Les personnes interrogées ont cependant cité **deux conditions** importantes pour elles, une fois ce service créé. La première est que le service reste mis à jour, pro-actif et prospectif. Cette démarche pourrait s'exprimer par la publicité du service auprès de dépositaires, la recherche permanente de données, le recensement d'informations non centralisées, ou encore par la digitalisation de sources plus anciennes. La deuxième condition mentionnée par les participants est que le service ne fasse pas de double emploi mais soit, au contraire,

---

<sup>1</sup> Cfr. partie 1.

complémentaire avec les données, les portails et les instituts de statistiques existant au niveau fédéral, régional ou européen (archives générales du Royaume, Statbel, IWEPS, ...).

Pour terminer, d'éventuels problèmes liés aux droits d'auteurs des données ont été mis en lumière par les participants. La possibilité d'archivage des recherches dépendant fortement des modalités de conventions conclues, il importerait dans l'idéal de créer un service d'archivage qui puisse centraliser des données, tout en s'adaptant aux exigences des différentes conventions avec les commanditaires de recherche.

## **3EME PARTIE : MISE EN PLACE DU SERVICE D'ARCHIVAGE**

### **Introduction**

---

Voici enfin la phase la plus importante de notre étude : offrir des propositions les plus concrètes possibles sur la manière de mettre en place un service d'archivage pour les sciences sociales en tenant compte du contexte particulier de la Belgique.

Dans ce rapport, les différentes facettes du futur service d'archivage ont été subdivisées en 7 parties :

- Ses missions
- Les types de données présentes dans le service
- Les procédures de dépôt des données
- Les procédures d'accès aux données
- Les méta-données et les aspects techniques du système
- Les modalités de création institutionnelles du service d'archives
- L'analyse de questions juridiques

Pour chacun de ces aspects, nous proposerons un certain nombre de scénarios possibles, non limitatifs. Parallèlement à ceux-ci, nous jetterons un regard sur les expériences étrangères et les avis des institutions, du monde académique et scientifique que nous avons recueillis et donnerons, enfin, pour résumer la situation, les points de consensus autour des différents scénarios imaginés. Nous consacrerons enfin, la dernière partie de ce chapitre aux aspects juridiques de la mise en place du portail.

Si certaines options peuvent être clairement tranchées, d'autres doivent encore faire l'objet d'une décision politique en haut lieu. Ce rapport mentionnera dès lors différentes alternatives étudiées, avec leurs forces et leurs faiblesses.

## **Chapitre 1. Un portail ou un service d'archivage ?**

---

### **1.1. QUESTIONS CENTRALES**

Le futur service créé doit-il se limiter à l'archivage de données en sciences sociales et à leur accessibilité via un portail, ou doit-il intégrer des missions supplémentaires (recherche et développement dans le domaine de l'archivage, publications, formations, envoi et conception de supports documentaires, ...) ? Autrement dit, quelles peuvent être les missions principales et secondaires du futur service d'archivage ?

## 1.2. SCENARIOS POSSIBLES

### 1.2.1 Missions principales du service d'archivage

- ☐ Garantir une conservation sécurisée des données
- ☐ Favoriser leur accès et leur réutilisation
- ☐ Faciliter un échange de données entre institutions et chercheurs
- ☐ Réaliser une certification garantissant/reconnaissant la qualité des données déposées
- ☐ .....

### 1.2.2 Missions secondaires du service d'archivage

- ☐ Activités de Recherche et Développement dans le domaine de l'archivage  
(intégration de la recherche à l'archivage)
  - Conférences
  - Séminaires
  - Publications
  - .....
- ☐ Formation aux techniques d'archivage
- ☐ Envoi de supports documentaires
- ☐ Coopération avec d'autres centres d'archive et des bases de données internationales
- ☐ .....

## 1.3. POSITIONNEMENTS

### 1.3.1 D'après les expériences étrangères

La mission principale défendue par les portails d'archive étrangers est de favoriser l'accès et la réutilisation des données. Cependant, tous les services d'archive que nous avons observés ne se limitent pas à l'archivage pur et simple de données, mais intègrent, au contraire, des missions supplémentaires. La **recherche et développement** dans le domaine de l'archivage (conférences, écoles d'été, séminaires, publications, ...) est ainsi une activité secondaire importante de nombreux services d'archivage, comprenant à la fois une optique d'étude et une optique de conservation et de valorisation des données. Les portails d'archive que nous avons examiné contiennent également de nombreuses **fonctionnalités périphériques** à une simple base de données documentaire (références de littérature, téléchargement possible d'outils pour manipuler ou analyser des données, liens vers d'autres banques de données documentaires, annonces pour des conférences, thesaurus, ...).

### 1.3.2 D'après les avis des institutions, du monde académique et scientifique

Pour le monde académique et scientifique, le service d'archivage pourrait constituer une importante 'mémoire collective'. La mission primaire de **favoriser l'accès et la réutilisation des données** est plébiscitée comme la mission essentielle du futur service d'archives. Celle-ci est suivie par une autre mission importante : favoriser **l'échange de données entre les**

**institutions et/ou les chercheurs.** Pour les personnes interrogées, il semble important que le service d'archivage ne se limite pas à être un simple portail statique. Au contraire, celles-ci souhaitent qu'il ait une dynamique proactive pour constamment rechercher, centraliser, mettre à jour des données et fournir de la documentation.

Ces deux **missions primaires** sont également plébiscitées par le monde institutionnel, avec en plus une mission de conservation des données. La démarche proactive que devrait avoir le service en matière de recherche de données, d'information et d'échanges internationaux est également mise en avant.

A propos des **missions secondaires**, l'optique des personnes interrogées est d'abord que le centre ne dépasse pas trop ses missions primaires pour ne pas faire de double emploi avec d'autres initiatives existantes (par exemple pour les activités de recherche et développement dans l'archivage telles que les conférences, les séminaires et les publications). Bien **documenter** des données et **coopérer** avec d'autres centres d'archives régionaux, ou communautaires, ou des organismes politiques, et internationaux sont cependant des missions secondaires importantes.

**Pour les deux groupes**, une certification de la qualité des données par le service d'archives est également intéressante, mais ce processus paraît difficile à mettre en place étant donné la variété des données et la multiplicité des critères existant pour vérifier et définir la notion de qualité.

Pour terminer, deux atouts paraissent intéressants pour ce futur service d'archives.

- le futur service d'archives pourrait s'occuper de former les utilisateurs aux **techniques d'archivage** et stimuler le **réemploi** de données déposées dans le portail.
- Le portail d'archive pourrait être doté de **différents outils** fonctionnels comme un système de newsletter ou de mailing-list avertissant les utilisateurs des différentes mises à jour dans le système, mais aussi d'un système de liens utiles vers des bases documentaires existantes. Des liens vers les contrats de dépôt et d'utilisation des données en vigueur, la législation ou encore vers des logiciels de détection du plagiat pourraient renforcer la confiance des dépositaires et des utilisateurs.

## 1.4. POINTS DE CONSENSUS

En résumé, plusieurs points de consensus ont été trouvés autour des missions du service d'archivage :

- ✓ Comme dans les services d'archive présents à l'étranger, le service doit **dépasser un simple portail d'archive** pour les acteurs rencontrés.
- ✓ **Deux missions principales** de ce service ont été mises en avant :
  - Favoriser l'accès et la réutilisation des données.
  - Décloisonner la recherche en facilitant un échange de données entre institutions et chercheurs et en mettant en relation la recherche internationale, nationale et régionale.
- ✓ **Deux missions secondaires** du service paraissent essentielles :
  - La production de supports documentaires.
  - La coopération avec d'autres centres d'archives et bases de données internationales.

- ✓ Dans un premier temps, le futur service doit commencer par un **minimum d'activités**, ne pas trop se disperser et se concentrer d'abord sur ses missions de base.
- ✓ Il est également important que le service soit **proactif** (faire de la publicité auprès des chercheurs, récolter des données, les mettre à jour, fournir de la documentation actualisée, ...), tout en veillant à ne **pas** faire de **concurrence ni double emploi** avec d'autres initiatives existant dans le domaine de l'archivage.

## **Chapitre 2. Types de données dans le service d'archivage**

---

### **2.1. QUESTIONS CENTRALES**

La deuxième question centrale porte sur les types de données les plus adéquates à insérer dans le futur service d'archivage. Par types de données, nous entendons les catégories de données, leurs disciplines, le caractère et l'origine des données, et enfin le mode de sélection de celles-ci.

### **2.2. SCENARIOS POSSIBLES**

#### **2.2.1 Catégories de données**

- ☐ Données quantitatives
- ☐ Données qualitatives
- ☐ Ressources multimédia
- ☐ Matériel non digital (supports sur papier)
- ☐ Autres types de données : .....

#### **2.2.2 Disciplines**

- ☐ Toutes les données se rapportant aux sciences sociales au sens large : sociologie, économie, psychologie, histoire, criminologie, démographie, statistique, autres, ...
- ☐ Uniquement la sociologie
- ☐ Uniquement des bases de données quantitatives, peu importe leur discipline
- ☐ .....

#### **2.2.3 Caractère des données**

- ☐ Données "brutes" (chiffres collectés, extraits d'interviews, ...)
- ☐ Données "travaillées" : rapports de recherche, articles, guides méthodologiques, questionnaires, déclaration de variables, plans de traitements, ... (documentation)
- ☐ Données personnelles
- ☐ Données anonymisées
- ☐ .....

#### **2.2.4 Mode de sélection des données publiées dans le service**

- ☐ Définition d'un concept de "sciences sociales" et typologie des données
- ☐ Bases de sélection des données liée à
  - la taille des données
  - la fiabilité de l'enquête
  - .....
  - quelles sont les bases de sélection possibles et souhaitées, ... ?

## **2. 3. POSITIONNEMENTS**

### **2.3.1 D'après les expériences étrangères**

Les types de données que nous avons observées dans les portails pour les sciences sociales à l'étranger sont majoritairement **quantitatives** et se rapportent aux **sciences humaines au sens large**. Nous avons constaté avec surprise qu'aucun des portails que nous avons ciblés n'avait défini ou même donné des références au concept de "sciences sociales". Ces services d'archive intègrent au contraire tous les types de données en sciences sociales dans un but d'**interopérabilité**, jugeant tout découpage dans les sciences humaine difficile et arbitraire.

**Les bases de sélection** des données dans ces expériences sont souvent leur intérêt pour la communauté scientifique à une large échelle en réponse à certains besoins et leur facilité de réutilisation.

Notons enfin que toutes les données présentes dans le portail d'archive sont toujours **anonymisées** et que des **normes standards** d'encodage sont utilisées pour encoder les données dans les bases documentaires. Les données présentes comportent toujours au moins de la **documentation** (fiches documentaires, abstracts, modèles de questionnaires, ...) et sont parfois accessibles de **manière brute** en s'enregistrant dans le portail.

### **2.3.2 D'après les avis des institutions, du monde académique et scientifique**

#### **K. CATEGORIES DE DONNEES**

Pour les représentants que nous avons rencontrés, **toutes les catégories de données** auraient a priori leur intérêt à figurer dans le portail. Il faut cependant veiller à ne pas faire double emploi avec des données déjà référencées ailleurs et il apparaît important de bien faire apparaître la spécificité d'un archivage digital par rapport à un archivage sur papier.

Leur optique se veut **réaliste** : il vaut mieux commencer par le minimum, c'est-à-dire de rendre d'abord disponible des **données quantitatives** (qui englobent beaucoup d'éléments et sont plus faciles à réutiliser), puis de s'attarder sur d'autres types de données, plus qualitatives, celles-ci étant souvent plus spécifiques.



## L. DISCIPLINES DES DONNEES ET MODE DE SELECTION

**Pour les disciplines des données et leur mode de sélection**, les personnes interrogées trouvent qu'une segmentation par discipline serait dangereuse. En effet, les distinctions de branches sont bien souvent arbitraires et difficiles à établir. Il faut donc prendre en compte **toutes les données** se rapportant aux sciences sociales et humaines.

Pour que le portail puisse être utile, il est par contre important de se poser certaines questions fondamentales sur ses priorités : à quels utilisateurs s'adresse t'il ? suivant quels objectifs ? et pour répondre à quels besoins ?

Il est également essentiel que la **qualité** et la **fiabilité** des données soit garantie. Pour se faire, sélectionner des fournisseurs de données en provenance du secteur public, du monde scientifique ou du monde académique peut être un moyen de garantir cette qualité.

## M. CARACTERE DES DONNEES

Ce sont des **données travaillées**, c'est-à-dire mises en forme et documentées, que les représentants que nous avons interrogés voudraient voir avant tout présentes dans le service. Pour accompagner les documents, une présentation des ressources sous forme de **fiches** contenant des méta-données<sup>1</sup> ou des variables standardisées faciliterait grandement leur lecture. Dans un deuxième temps, le service pourrait contenir des données brutes, c'est-à-dire non mises en forme (chiffres, entretiens ou extraits d'entretiens). Ces données pourraient être réutilisées par d'autres chercheurs s'intéressant aux mêmes thématiques de recherche, malgré le fait que les données qualitatives recueillies soient souvent personnelles, très spécifiques et peu réutilisables. Pour rester dans un cadre légal, il apparaît également essentiel que les données accessibles dans le service d'archivage soient **anonymisées**<sup>2</sup>.

## 2.4. POINTS DE CONSENSUS

Les positions des personnes rencontrées sur le type de données à mettre dans un service d'archivage pour les sciences sociales en Belgique sont assez unanimes. Le service devrait se développer de manière réaliste **en deux temps**.

- ✓ Dans un premier temps, d'une manière limitative, c'est-à-dire en se concentrant sur les données et les ressources quantitatives des sciences sociales et leur documentation sous forme de fiches méthodologiques incorporant les méta-données.
- ✓ Dans un deuxième temps, en **élargissant** la cible à d'autres types de données plus qualitatives.

Les données seraient anonymisées et la présentation et l'encodage des ressources devront être effectués d'une manière standard.

---

<sup>1</sup> Nous reviendrons sur les méta-données dans le chapitre V. : Méta-données et aspects techniques du système.

<sup>2</sup> Cfr. l'analyse juridique ci-dessous.

## Chapitre 3. Dépôt de données dans le service d'archivage

---

### 3.1. QUESTIONS CENTRALES

Un service d'archivage ne pourrait pas exister sans une alimentation régulière de celui-ci. Qui seront les fournisseurs de données dans le système ? Uniquement des fournisseurs officiels de données ou aussi d'autres fournisseurs ? Notre préoccupation centrale sera également de savoir comment attirer des dépositaires, qu'ils soient des centres de recherche ou des particuliers, à venir déposer des données dans le service d'archives.

### 3.2. SCENARIOS POSSIBLES

#### 3.2.1 *Dépositaires de données dans le portail*

- ☐ Agences de statistique officielles nationales : INS, IWEPS, ...
- ☐ Agences de statistique officielles internationales : eurobarometer...
- ☐ Universités, FNRS, enseignement supérieur
- ☐ Administrations et politiques
- ☐ Sociétés privées : SONECOM, ...
- ☐ Associations
- ☐ Presse et médias
- ☐ Etudiants
- ☐ Toute autre personne
- ☐ .....

#### 3.2.2 *Comment alimenter le système et convaincre les centres de recherche de déposer leurs données ? Comment attirer les chercheurs ?*

- ☐ Chaque organisme ou chercheur qui fournit des données dans le portail ou s'y implique peut accéder, en contrepartie, aux données du service d'archives.
- ☐ Chaque organisme ou chercheur qui fournit des données dans le portail ou s'y implique peut recevoir gratuitement, en contrepartie, les données du portail sur des formats cd par exemple.
- ☐ Tous les centres de recherches, ou les chercheurs qui ont des subventions des organismes subventionnant le service d'archivage (l'Etat fédéral, les communautés, les régions, ...) doivent, en contrepartie, mettre leurs données en ligne sur ce portail.
- ☐ Les chercheurs et organismes qui publient dans le portail d'archive, sont en compensation, assurés de recevoir une "publicité" à leur recherche et leurs données...
- ☐ .....

### 3.3. POSITIONNEMENTS

#### 3.3.1 D'après les expériences étrangères

Les données déposées dans les services d'archivage proviennent surtout d'**agences de statistiques nationales et internationales**, de **fournisseurs** issus de la communauté scientifique (universités ou chercheurs isolés ...) ou des sources historiques. Très peu de données commerciales sont proposées. Pouvant être utilisées à des fins de marketing, celles-ci ne sont pas du tout considérées comme prioritaires dans un service d'archivage, et la qualité scientifique de leur collecte et de leur analyse est souvent mise en doute. Les données sont soit **achetées** auprès de grands fournisseurs comme des agences de statistique, soit **déposées** par des chercheurs sur base de leur libre initiative ou pour répondre aux exigences d'un financement reçu. L'alimentation du système se fait généralement par les travailleurs des centres d'archive eux-même, après avoir conclu des **conventions** avec les fournisseurs sur les droits d'auteurs et la réutilisation des données (licence agreement...). Enfin, certains portails d'archive comme le DANS possèdent également une fonction d' « **easy data depositing** », dans lequel un chercheur isolé peut s'identifier et lui-même déposer des données qu'il souhaite mettre à disposition.

#### 3.3.2 D'après les avis des institutions, du monde académique et scientifique

##### A. DEPOSITAIRES DES DONNEES

Les acteurs que nous avons rencontrés sont en faveur de **plusieurs types de dépositaires** dans le service d'archivage. Leur préférence semble avant tout marquée pour les bases de données publiques belges et étrangères, c'est-à-dire pour les quatre premières catégories proposées : les agences de statistiques officielles nationales, les agences de statistiques officielles internationales, les données des administrations ou du politique et des données en provenance des universités et de l'enseignement supérieur. Les **données fournies par le secteur public** sont donc essentiellement valorisées, d'autres types de données pouvant venir se greffer par la suite, sur base volontaire, dans le service d'archives. Les quatre premières catégories de dépositaires sont en tous cas perçus par nos représentants du monde académique et scientifique comme fournissant de données fiables et de bonne **qualité méthodologique**. Les acteurs sont plus méfiants pour des données issues de sociétés privées qui pourraient y être déposées à des fins commerciales. Ces types de données nécessitent donc qu'un certain cadre légal soit défini.

Une fois de plus, le service d'archives aurait toutes ses raisons d'exister pour référencer ce type de données, mais de nouveau, à condition de ne **pas** faire de **double-emploi** ni de concurrence avec les statistiques officielles ou les bases de données déjà référencées<sup>3</sup>.

##### B. ALIMENTATION DU SYSTEME

Comment alimenter le système et convaincre les dépositaires (centres de recherche, agences de statistiques, administrations et politiques,...) de déposer leurs données est une question cruciale pour la mise en place de ce service d'archivage. Les acteurs que nous avons

---

<sup>3</sup> Vous trouverez en bibliographie de ce rapport un référencement de quelques sites Internet de statistiques et de bases de données.

rencontrés ont tous insistés pour que le **processus d'alimentation** soit **gagnant-gagnant**, c'est-à-dire bénéfique, tant pour les dépositaires de données, que pour coordinateurs du service d'archives et les utilisateurs qui y consulteront des données. Des divergences sont cependant présentes dans la manière de concrétiser ce processus. En effet, plusieurs idées sont proposées :

- a. Une première **optique** serait **une alimentation basée sur la libre initiative des dépositaires**. Pouvoir valoriser et faire connaître les résultats de ses recherches est très fréquemment un problème pour un chercheur qui ressent bien souvent des frustrations à cet égard. Face à cette situation, une solution simple existerait : publier, dans une interface centralisée et fréquentée, les résultats de leurs propres recherches en ligne, ce qui assurerait aux dépositaires une visibilité et une valorisation de leurs travaux.
- b. La **deuxième optique** imaginée serait **plus coercitive**. Les chercheurs ou les agences de statistiques seraient obligées de publier leurs résultats de recherche dans le système : en envoyant leurs données, leurs résultats de recherche et/ou leur publication au service d'archives qui les publierait ensuite dans un portail accessible aux utilisateurs. Cette obligation pourrait se passer de trois manières :
  - c. Soit en adoptant juridiquement une loi sur la valorisation de données dans le cadre de recherches
  - d. Soit en prévoyant dans les contrats de recherche chercheur une mise à disposition publique et standardisée de ses recherches<sup>4</sup> sur le portail. Un contrôle pourrait s'effectuer par les organismes finançant les recherches (communautés, régions, universités, ...).
  - e. Soit, en prévoyant, dans les contrats de recherche dont le financement se fait par tranches une dernière tranche de financement qui ne serait acquise qu'après le dépôt de données dans le portail.
- c. **La troisième piste des contreparties** propose que chaque organisme ou chercheur qui fournit des données dans le portail ou s'y implique peut accéder, en contrepartie, aux données du service d'archives ou recevoir, en contrepartie, les données du portail sur des formats et recueil moins d'adeptes. Celle-ci semble cependant plus compliquée à mettre en place.

Les deux pistes principales ont chacune **leurs avantages et leurs inconvénients**.

Si l'optique de la libre initiative paraît assez facile à mettre en œuvre, les acteurs du monde de la recherche rencontrés ont généralement déploré un **manque de temps et de moyens accordé dans les conventions de recherche pour pouvoir valoriser leurs données**. Cette optique paraît donc assez utopique si elle n'est pas accompagnée de politiques incitatives à la valorisation de la recherche, c'est-à-dire, notamment, par des budgets propres à cette valorisation dans les conventions de recherche. Actuellement, cette valorisation est en souvent liée à la bonne volonté du chercheur au-delà de son temps de travail ! Il serait donc intéressant de remédier à ce problème pour inciter les chercheurs à un meilleur archivage des données.

L'optique de l'obligation, en revanche, rendrait l'archivage des sciences beaucoup plus systématique. Cependant, celle-ci nécessiterait que deux conditions importantes soient

---

<sup>4</sup> Aujourd'hui, des exigences pour la conservation ou la mise à disposition de données sont présentes dans certaines conventions de recherche, mais celles-ci sont loin d'être très répandues et présentes dans toutes les conventions.

remplies. D'un côté, que d'importants moyens soient investis dans des dispositifs de **contrôle** de cet archivage. D'un autre côté, que des **accords** institutionnels pour une conservation et une valorisation centralisée de la recherche soient trouvés entre les différents organismes finançant les recherches (régions, communautés, FWO, FNRS, politique scientifique fédérale, ...), les gestionnaires du portail d'archive et les dépositaires de données. Ces accords institutionnels devraient également fournir aux dépositaires de données du temps et/ou des moyens pour la valorisation de leurs données dans le portail d'archives.

Au niveau des bases de données susceptibles d'être mises à disposition du service d'archivage, **plusieurs bases de données existent**. Comme plusieurs fournisseurs de données du secteur public (INS, SPF Sécurité sociale, ...) possèdent déjà leur propre base de données et leur propre système d'accès, les bases de données existantes pourraient être fusionnées avec celles du service d'archivage pour autant que des accords ou des synergies soient trouvés entre les gestionnaires du portail d'archive et les gestionnaires de ces bases de données<sup>5</sup>. Une formule alternative serait de mettre un système de lien du portail d'archive vers les bases de données publiques difficiles à importer.

#### 4. POINTS DE CONSENSUS

Afin d'obtenir des données de bonne **qualité** dans le portail, les positions des acteurs semblent assez unanimes pour que les dépositaires soient primitivement issus du secteur public.

- ✓ Dans un premier temps, le système proposerait des données issues de bases de données publiques officielles belges et étrangères (milieu de la recherche, d'administrations publiques ou d'agences de statistiques nationales ou internationales). Ces données seraient disponibles en fonction des conventions et des accords institutionnels conclus, soit directement, soit par un système de liens.
- ✓ Dans un deuxième temps, le système pourrait être élargi à d'autres types de données plus qualitatives en provenance de fournisseurs plus diversifiés, à condition que la qualité de leurs ressources soit garantie et que celles-ci ne soient pas déposées dans le portail à des fins commerciales.
- ✓ Un cadre légal et des conventions de dépôt et de réutilisation des données, devront bien évidemment être définis<sup>6</sup>.

Le mode d'alimentation des données dans le portail d'archive devrait être gagnant à la fois pour les gestionnaires du portail, pour les dépositaires des données et pour les utilisateurs de ce portail. Il pourrait

- ✓ être alimenté à l'initiative des dépositaires à condition qu'il y ait des **politiques incitatives** à la valorisation de la recherche, c'est-à-dire du temps et des moyens accordés dans les conventions de recherche pour cette valorisation.
- ✓ être un processus obligatoire pour les dépositaires à condition qu'il y existe des **dispositifs de contrôle** et des **accords institutionnels préalables entre les organismes finançant la recherche, les gestionnaire du portail et les dépositaires de données.**

<sup>5</sup> Vous trouverez en bibliographie de ce rapport un référencement de quelques sites Internet de bases de données en sciences sociales.

<sup>6</sup> Cet aspect sera abordé dans le Chapitre 7 : Modalités juridique du dépôt et de la réutilisation des données issues de la recherche.

## **Chapitre 4. Accès et dynamique de gestion des données**

---

### **4.1. QUESTIONS CENTRALES**

Une question importante dans la mise en place du service d'archivage est son public cible. Pour quel types d'utilisateurs ce service d'archivage doit-il être conçu ? Uniquement pour des spécialistes des sciences sociales ou également pour des profanes en la matière ? De plus, comment sera accessible le portail d'archive et quelles seront les procédures d'accès et de réutilisation des données ? Ce sont ces interrogations importantes que cette partie va examiner.

### **4.2. SCENARIOS POSSIBLES**

#### ***4.2.1 Utilisateurs potentiels des données***

- ☐ Chercheurs
- ☐ Administrations de la recherche
- ☐ Monde politique et les administrations publiques
- ☐ Entreprises
- ☐ Presse et médias
- ☐ Etudiants
- ☐ Toute autre personne
- ☐ .....

#### ***4.2.2 Procédure d'accès au système***

##### **C. SITUATION**

- ☐ Portail accessible sur Internet
- ☐ Portail accessible sur Intranet
- ☐ Portail accessible des deux manières

##### **D. ACCES AU CONTENU**

- ☐ Accès directement aux données (virtual library)
- ☐ Système de liens vers des données existant sur d'autres sites (portail d'information)
- ☐ Système de référencement de données (référothèque)
- ☐ .....

##### **E. MOYEN D'ACCES AUX DONNEES**

- ☐ Accès libre et direct aux données

- ☐ Accès aux données avec identification, login et mot de passe
- ☐ Accès aux données sur base d'un formulaire à remplir

#### F. DANS LE CAS D'UN ACCES DIRECT AUX DONNEES

- ☐ Accès nécessitant de s'enregistrer pour accéder à l'ensemble du portail d'archivage
- ☐ Accès ouvert uniquement aux fiches documentaires des ressources et nécessité de s'enregistrer dans le système pour avoir accès aux données
- ☐ Accès ouvert pour les fiches documentaires et les documents travaillés, mais fermé pour les données brutes
- ☐ Accès totalement ouvert pour l'ensemble du portail d'archivage
- ☐ Fixation de l'accès aux données par le dépositaire du document dans le portail selon des catégories prédéfinies : ex : données accessibles à tout le monde, données uniquement réservées à l'enseignement et à la recherche, ...
- ☐ .....

#### G. ACCES PAYANT OU GRATUIT

- ☐ Accès gratuit aux données
- ☐ Accès payant aux données
- ☐ Formule combinée
- ☐ .....

### 4.2.3 Dynamique du portail d'archive ?

- ☐ Différents rôles et statuts possibles autour de l'application : lecteur, rédacteur, validateur, administrateur, webmaster, ...
- ☐ Gestion exclusive de l'application informatisée par le ou les organismes gérant le service d'archivage
- ☐ Gestion du portail d'archive par un ensemble d'institutions partenaires travaillant ou s'intéressant au domaine de l'archivage (comité de gestion)
- ☐ .....

## 4.3. POSITIONNEMENTS

### 4.3.1 D'après les expériences étrangères

La **communauté scientifique**, le **secteur de l'enseignement** et les **administrations** constituent généralement le **public cible** des portails d'archive présents à l'étranger. Des accès aux rubriques du portail et à de la documentation sur des ressources sont généralement possibles sans s'enregistrer, mais un enregistrement sécurisé ou non est nécessaire pour accéder aux données brutes ou à de la documentation de certaines données plus sensibles. Les **données** présentes peuvent être directement accessibles, uniquement référencées ou encore soit commandables sur certains types de supports comme des CD-rom qui sont généralement payants. Dans tous les cas, les utilisateurs du portail devront prendre connaissance des **conditions de réutilisation** des données qu'ils vont employer.

Les **catégories d'accès** aux données sont habituellement fixées par les dépositaires de données parmi des niveaux d'accès existants (données accessibles à tout le monde, données accessibles uniquement au secteur de la recherche et de l'enseignement, données accessibles mais non publiables, ...)

#### **4.3.2 D'après les avis des institutions, du monde académique et scientifique**

##### **H. UTILISATEURS POTENTIELS DES DONNEES**

Les participants rencontrés s'opposent généralement à une limitation de l'accès du service aux seuls déposants, ce qui serait trop restrictif et fermerait beaucoup de portes. Au contraire, cet accès pourrait être **ouvert** prioritairement au monde académique et scientifique, ... et plus tard, élargi à d'autres publics comme les administrations ou les étudiants. Un bémol est cependant posé pour l'accès au service des sociétés privées qui pourraient utiliser celles-ci à des fins commerciales.

##### **I. PROCEDURE D'ACCES AU PORTAIL**

Les représentants du monde de la recherche interrogés sont en faveur d'un portail avec un **accès gratuit** aux données. Si certains fournisseurs de données voudraient fournir des données payantes, il appartiendrait au final au service d'archives de les mettre à disposition gratuitement de ses utilisateurs. Les acteurs rencontrés sont d'accord de payer pour certains frais matériels comme l'envoi de supports documentaires sur CD par exemple, mais pas de couvrir des frais de fonctionnement du service en lui-même qui devrait être géré par des fonds publics.

Les acteurs plébiscitent également que le portail soit **disponible sur Internet** (plutôt que sur un intranet), beaucoup d'utilisateurs pouvant y accéder facilement.

Dans une optique d'« **open acces** » aux données, les acteurs interrogés sont en faveur d'un accès facile et ouvert aux ressources déposées, à l'exception de conditions spéciales. De manière concrète, une procédure d'accès aux données à **deux niveaux** est la plus souhaitable. **Voici ce que nous proposons :**

- Sans s'enregistrer, les utilisateurs du portail pourraient accéder à un descriptif du service d'archives, avoir un aperçu des principales ressources déposées dans le portail faire des recherches sur les ressources du portail et visualiser les fiches descriptives de chaque ressource comportant ses principales méta-données<sup>7</sup> et un résumé du contenu.
- En s'identifiant au système, les utilisateurs pourraient accéder aux ressources proprement dites, c'est-à-dire aux données brutes récoltées (chiffres, transcriptions d'interviews, ...) et à de la documentation sur ces données (rapports de recherche, explication de la méthodologie utilisée, grille d'analyse des données, questionnaires employés, ...). Il pourrait également être de décider de pouvoir accéder uniquement à une partie du contenu<sup>8</sup>. Les utilisateurs pourraient également sauvegarder ou

<sup>7</sup> Des méta-données comme le titre, le(s) contributeur(s), la langue de la ressource, la date de sa création et un petit résumé de celle-ci peuvent apparaître sur la fiche descriptive de chaque ressource.

<sup>8</sup> Pour l'accès au contenu des ressources, une autre solution que nous avons proposée dans les scénarios ci-dessous serait que chaque dépositaire de documents fixe des conditions d'accès selon des



commander des données, déposer des données sur le portail (pour les organismes habilités à publier ces résultats) et avoir un aperçu des utilisateurs et des fournisseurs de données dans le portail<sup>9</sup>.

- Tous les utilisateurs du portail venant télécharger une ressource seraient obligés de signer une **convention de dépôt** (pour les dépositaires) **ou de réutilisation des données** (pour les utilisateurs). Ces conventions contiendraient un certain nombre d'obligations légales et d'exigences en matière de dépôt et de réemploi de données déjà publiées (obligation de citation de la source, respect de la vie privée,...). Les différents sites Internet de bases de données ou de portails d'archives que nous mentionnons tout au long de ce rapport<sup>10</sup> peuvent, en outre, servir d'exemples de modèles d'accès à respecter. Nous aborderons cet aspect dans la partie juridique du rapport.

## J. DYNAMIQUE DU PORTAIL D'ARCHIVE

Pour les acteurs interrogés, il est essentiel que le portail d'archive s'actualise et reste dynamique. La dynamique de gestion qui sera choisie dépendra essentiellement des modalités de gestion du service d'archives et des implications institutionnelles dans le portail (implication des archives générales du royaume, des communautés, des régions, ...). Nous aborderons cet aspect dans le dernier chapitre de ce rapport : "Modalités institutionnelles de création du service d'archives".

Traditionnellement, **5 rôles différents**<sup>11</sup> peuvent être distingués autour d'une application informatisée. Ces rôles recouvrent des droits différents d'accès au cœur même de l'application et pourraient être employés pour le futur service d'archives belge.

- **Lecteur** : il a la possibilité de consulter le contenu de la base de données via l'interface
- **Rédacteur** : il a le droit de rédiger des fiches à intégrer (via procédure) dans la base de données et de modifier celles qu'il a déjà rédigées.
- **Valideur** : il valide et a le droit d'éditer les fiches des rédacteurs qu'il supervise dans l'application (activation) ou d'en retirer (désactivation).
- **Administrateur** : il gère les accès et possède l'ensemble des droits sur la base de données : modification de fiches, suppression, modification de la structure de la base

---

catégories prédéfinies : ex : contenu accessible à tout le monde, contenu accessible partiellement ou accessible par enregistrement, ... L'accès aux données et à la documentation devrait cependant rester **dans le cadre légal fixé par les conventions de recherche**.

<sup>9</sup> Le site Internet d'information juridique <http://www.strada.be> est à cet égard un portail d'accès aux données structuré de manière intéressante.

<sup>10</sup> Vous trouverez en bibliographie de ce rapport un référencement de quelques sites Internet de bases de données en sciences sociales.

<sup>11</sup> Claire LOBET-MARIS, Véronique DUMONT, Véronique LAURENT, Evelien DE PAUW, Paul PONSAERS, *Une bibliothèque virtuelle pour la Police Fédérale : analyse de la faisabilité du projet et cahier des charges méthodologique de l'application*, Rapport de recherche du projet "Agora-Bibliothèque Virtuelle", 2005, p126. (Version électronique disponible sur [http://www.fundp.ac.be/recherche/publications/page\\_view/56910/](http://www.fundp.ac.be/recherche/publications/page_view/56910/))

de données (ajout de champs, ajout /suppression d'items préencodés), modification du statut d'une personne.

- **Webmaster** : il gère le design et l'interface de l'application, l'ergonomie, la mise en ligne, les liens entre l'application sur Intranet et Internet (copie, mise à jour).

#### 4.4. POINTS DE CONSENSUS

Les points de consensus autour de l'accès au service d'archivage des données sont les suivants :

- ✓ Les ressources du portail d'archive devraient être accessible gratuitement et sur Internet.
- ✓ Le public cible du service doit aller plus loin que les dépositaires de données eux-même. Les utilisateurs prioritaires sont avant tout les représentants du monde de la recherche et du monde académique.
- ✓ Le portail doit être le plus facile et le plus ouvert possible (dans la limite de ce que permettent les conventions de recherche, les contrats de dépôt et les contrats de réutilisation des données).
- ✓ Un système d'accès à deux niveaux (avec et sans enregistrement) est souhaitable.
- ✓ Des conventions de dépôt et de réutilisation des données par les dépositaires et les utilisateurs doivent être signées.
- ✓ La dynamique de gestion du système est fortement dépendante des implications institutionnelles dans le portail.

### Hoofdstuk 5. Technische voorwaarden voor archivering

Voor de langetermijnarchivering van digitale archiefdocumenten is een digitaal archiefdepot nodig. Een digitaal depot is niet zomaar een technische infrastructuur waarbinnen digitale data worden beheerd, maar maakt deel uit van een digitaal archiveringssysteem en krijgt dus mee vorm door de visie op langetermijnarchivering, de kwaliteitsvereisten, de normen en de procedures die hiervoor worden gehanteerd. Het digitaal archiveringssysteem moet ervoor zorgen dat digitale data op het tijdstip van raadpleging leesbaar, begrijpbaar voor mens en computer, bruikbaar en geloofwaardig zijn. Het digitaal archiefdepot geeft dit archiveringssysteem mee gestalte.

De kern van een digitaal archiefdepot wordt gevormd door drie processen, zijnde:

1. de opname en verwerking van de digitale data en hun metadata in het digitaal archiefdepot (input)
2. de lange termijn opslag en het beheer van de digitale data en hun metadata
3. het beschikbaar stellen van de digitale data en hun metadata die in het digitaal archiefdepot worden beheerd (output)

## 5.1. PROBLEMATIEK VAN DIGITALE DUURZAAMHEID

Een van de onderzoeksvragen betreft de problematiek hoe digitale archiefdocumenten op een duurzame en betrouwbare wijze worden gearcheeerd. De lange termijnarchivering van digitale data is omwille van diverse redenen een uitdaging met veel drempels. Deze worden hieronder één voor één opgesomd zodat duidelijk is welke oplossingen nodig zijn. Ambtenaren, IT-verantwoordelijken, documentbeheerders, onderzoekers en archivariissen worden alsmaar meer geconfronteerd met het bewaren en het archiveren van digitale archiefdocumenten. Digitaal archiveren is niet vanzelfsprekend, maar vraagt bijzondere oplossingen.

### 5.1.1. DE TECHNOLOGISCHE VEROUDERING

Voor het bekijken van digitale data is steeds een bepaalde hard- en softwareconfiguratie nodig. Aangezien men dient uit te gaan van het principe dat gearcheerde data een langere levensduur hebben dan de hard- en software omgeving waarbinnen ze werden gecreëerd of beheerd, dient men een oplossing te hebben voor de technologische veroudering. Een digitaal bestand kan immers een lange of zelfs permanente bewaartermijn hebben terwijl de gemiddelde IT-infrastructuur slechts een levensduur van gemiddeld 5 tot 10 jaar heeft. Digitale media zoals harde schijven, CR-R's en tapes hebben een kortere levenstermijn dan traditionele dragers zoals perkament, papier en microfilm.

### 5.1.2. DE GROTE HOEVEELHEID DOCUMENTEN

De administraties, onderzoeksinstituten en onderzoekers maken volop gebruik van IT voor het aanmaken en uitwisselen van documenten. De hoeveelheid digitale data neemt elke dag toe. Zelfs wanneer archiefwaardering en selectie goed wordt toegepast, zullen archiefdiensten met een grote toestroom aan digitale data worden geconfronteerd. Hiervoor zullen passende oplossingen zoals geautomatiseerde archiefbewerkingen en bulkverwerkingen nodig zijn. Deze processen dienen echter nauwlettend gecontroleerd te worden. Strengere kwaliteitscontroles en foutopsporings- en foutverbeteringsmechanismen zullen noodzakelijk zijn.

### 5.1.3. DE ARCHIEFWAARDERING EN SELECTIE

Digitale data nemen fysiek nauwelijks plaats in. Men kan zich bijgevolg de vraag stellen of archiefwaardering en selectie nog wel nodig zijn en of niet alle digitale data kunnen gearcheerd worden. Opslag wordt immers almaar goedkoper. Toch blijven archiefwaardering en selectie noodzakelijk. Goed archiefbeheer vraagt dat data die geen waarde meer hebben, worden vernietigd. Digitaal archiveren is immers een complex probleem dat veel onderzoek, tijd en middelen vraagt. Die worden bij voorkeur hoofdzakelijk aangewend voor data met een archiefwaarde. Het is zinloos om data zonder archiefwaarde te bewaren, bijkomende handelingen van gebruikers te vragen of om bijzondere vereisten te stellen aan informatiesystemen waarbinnen geen data worden geproduceerd. Immers, het creëren van goed archiveerbare digitale data van een hoge kwaliteit, het intellectueel beheer en het leesbaar houden vragen, in tegenstelling tot de opslag, wel veel middelen en inspanningen. Archiefwaardering is de sleutel om de data van complexe en technologie-afhankelijke systemen te archiveren. Archiefwaardering speelt ook mee in de keuze van een bepaald bestandsformaat als archiveringsformaat. Archiefwaardering en selectie maken efficiënter beheer mogelijk. Tenslotte houdt selectie ook de functionele vereisten voor de infrastructuur van het archiveringssysteem in de hand.

#### 5.1.4. DE VERSCEIDENHEID

De digitale data die momenteel gecreëerd en ontvangen worden zijn van een heel verscheiden aard. Er zijn niet alleen de diverse types digitale objecten (tekstverwerkingsbestanden, spreadsheets, e-mails, databanken, afbeeldingen, audio-visuele materialen, websites, GIS, CAD, virtuele modellen, enz.) ook de hard- en softwareconfiguraties zijn heel verscheiden. Voor elk digitaal bestand is een passende archiveringsoplossing nodig. Rekening houdend met de grote verscheidenheid aan besturingssystemen en applicaties is dit geen evidentie.

#### 5.1.5. DE AUTHENTICITEIT EN BETROUWBAARHEID

Digitale data bieden het voordeel dat ze na vastlegging nog steeds bewerkt kunnen worden. Digitale data kunnen snel worden aangepast. In veel gevallen kan de wijziging achteraf niet waargenomen worden. Hierdoor kunnen twijfels rond de betrouwbaarheid rijzen zodat passende maatregelen vereist zijn. De archivaris dient ervoor te zorgen dat de digitale bestanden niet onrechtmatig worden gewijzigd en dat eventuele manipulaties opgespoord en ongedaan worden gemaakt. Alleen zo is hij/zij in staat om de geloofwaardigheid van de gearcheerde data te verzekeren.

#### 5.1.6. DE ARCHIVERING VAN DE CONTEXT

Digitale bestanden zijn in de toekomst slechts bruikbaar wanneer ze door de gebruiker geïnterpreteerd kunnen worden. Met andere woorden, de gebruiker van de data moet weten binnen welke context de data werden gecreëerd of ontvangen, wat de functie van de data was en welke betekenis het bestand inhoudt. Hij/zij moet daarvoor minimaal weten binnen welk werkproces / onderzoek de data werden gecreëerd. Dit contact dreigt in de digitale omgeving verloren te gaan.

#### 5.1.7. DE ONTSLUITING EN HET TOEGANKELIJK MAKEN

Om digitale data snel opspoorbaar te maken, moet de informatie over de context aan de archiefgebruiker worden gecommuniceerd zodat hij/zij de data ten volle kan begrijpen. Interpretatie van de digitale bestanden is slechts mogelijk wanneer ze leesbaar zijn en er bijgevolg een oplossing is voor het digitale duurzaamheidsprobleem.

## 5.2. BEWAARSTRATEGIEËN

#### 5.2.1. HARD COPY

In de hard copy strategie wordt het digitaal bestand overgezet naar microfilm of afgedrukt op papier.

De archiefwetenschap gaat echter uit van de archivering van data in hun oorspronkelijke, primaire vorm: wat digitaal (op papier) ontstaat, wordt digitaal (op papier) gearcheerd. Bij omzetting naar papier of microfilm gaat een essentiële karakteristiek van het bestand ("het digitaal-zijn") verloren. Alleen al vanwege deze reden wordt de hard copy strategie beter niet toegepast. Bovendien spelen bij overzetting naar papier of microfilm nog andere factoren mee:

- de data verliezen hun 'digitale voordelen' zoals herbruikbaarheid, centrale bewaring en decentrale terbeschikkingstelling, geautomatiseerde archiefbeschrijvingen en zoekopdrachten, enz.

- functionaliteiten of bepaald gedrag van het bestand gaan verloren
- voor de vernietiging en de vervanging van data bij overheidsinstanties is de goedkeuring van de Algemene Rijksarchivaris of diens gemachtigde vereist (art. 5 Archiefwet 24 juni 1955)
- moeilijk te vermijden dat de digitale versies nog worden gebruikt als basis voor de handelingen: de vertrouwdsheid met digitale informatie groeit en de digitale versies zal men in werkprocessen als de primaire kopie blijven beschouwen én gebruiken.
- niet alle essentiële informatie wordt altijd afgedrukt
- niet alle digitale bestanden kunnen gemakkelijk naar papier of microfilm worden overgezet (bijv. GIS, CAD, multimedia objecten, databanken)
- hogere kostprijs: overzetten naar papier en microfilm is duurder dan digitale archivering.

Afdrukken op papier of overzetten op microfilm kan in principe enkel als tijdelijke archiveringsoplossing worden toegepast, in afwachting van een volwaardige digitale archiveringsprocedure. Deze optie is overigens niet toepasbaar voor alle types digitale documenten; enkel digitale bestanden met een papieren equivalent kunnen gemakkelijk afgedrukt. Een belangrijke vereiste is dat alle essentiële informatie mee wordt afgedrukt op papier of microfilm.

## 5.2.2. BEWAREN VAN DE TECHNOLOGIE

### A. *COMPUTERMUSEUMSTRATEGIE*

Deze piste bestaat uit het bewaren van de originele hard- en software waarmee de digitale bestanden werden gecreëerd. Op die wijze wordt een oude computerconfiguratie in stand gehouden zodat de computerbestanden in hun oorspronkelijke vorm raadpleegbaar blijven.

Voor middellange- en langetermijnbewaring is deze oplossing niet praktisch of haalbaar:

- alle verschillende configuraties moeten bewaard worden
- hard- en software hebben een beperkte levensduur
- oude hardware onderdelen worden almaar schaarser
- IT-kennis die nodig is voor het werken met oude hard- en software verdwijnt
- productondersteuning is meer mogelijk na verloop van tijd
- overzetten van digitale data naar nieuwe dragers is noodzakelijk vanwege de degradatie van de dragers. De kans is klein dat nieuwe apparaten en bijhorende stuurprogramma's op oude computers kunnen worden geïnstalleerd .

Deze piste is slechts haalbaar voor de korte termijnbewaring (5 à 10 jaar) van digitale archiefdocumenten. De museumstrategie is bijgevolg maar bruikbaar voor de bewaring van bestanden waarvan de bewaartermijn niet langer is dan die van de levensduur van de technologie of als tijdelijke oplossing in afwachting van een duurzamere archiveringsoplossing. Oude computerconfiguraties kunnen soms nog gebruikt worden voor het recupereren van data in verouderde formaten.

### B. *EMULATIE*

Bij emulatie wordt niet de originele hard- en software bewaard maar wordt het vereiste platform op een toekomstige computerconfiguratie gereconstrueerd zodat de computerbestanden in hun oorspronkelijk formaat raadpleegbaar zijn.

Emulatie kan op diverse niveau's worden toegepast. Men kan computerhardware, besturingssystemen, specifieke software of een combinatie van dit alles nabootsen. Emulatie is mogelijk op basis van configureerbare chips (emulatie door hardware) of op basis van computerprogramma's (emulatie door software).

Inmiddels bestaan verschillende visies op de wijze waarop emulatie voor digitale archivering kan worden toegepast:

- Jeff Rothenberg: Emulation Virtual Machine<sup>12</sup>
- Steve Gilheany: Turing Machine<sup>13</sup>
- Raymond Lorie: Universal Virtual Machine (data preservatie, programma preservatie)<sup>14</sup>
- Cedars & Camileon project: Migration on request<sup>15</sup>

Emulatie heeft een aantal interessante voordelen:

- in theorie kunnen de bestanden in hun oorspronkelijk formaat gearchiveerd worden:
  - alle originele eigenschappen en functionaliteiten blijven behouden
  - er gaan geen elementen verloren ten gevolge van omzettingen
  - de authenticiteit van de digitale data is gemakkelijker te garanderen
- de formaten waarin de data zijn opgeslagen, hoeven niet omgezet te worden telkens de formaten in onbruik raken
- de kostprijs is niet afhankelijk van het aantal gearchiveerde digitale bestanden.

Anderzijds zijn er ook een aantal nadelen aan emulatie verbonden:

- emulatie is technisch complex: de nodige know-how en expertise voor ontwikkeling en onderhoud zijn in archieven niet aanwezig. Archieven zijn bijgevolg afhankelijk van externe diensten en partners. Dit botst met de doelstelling om een zelf voorzienig digitaal archief te bouwen.
- emulatie heeft hoge ontwikkelings- en onderhoudskosten: kunnen archieven die nu voor deze benadering kiezen de financiële inspanningen in de toekomst blijven leveren?
- de platformen waarop emulatieprogramma's draaien, evolueren, wat op termijn omzettingen van of aanpassingen aan emulatieprogramma's zal vergen
- overkill: bepaalde emulatiebenaderingen gaan uit van de volledige reconstructie van de oorspronkelijke applicaties met alle functionaliteiten, terwijl in principe een viewer volstaat om de data weer te geven. Emulatie richt zich hoofdzakelijk op de lange termijn bewaring van systemen en software, terwijl de archivaris in eerste plaats de archivering van het digitaal bestand beoogt
- de bescherming van het auteursrecht op hard- en software houdt restricties in voor reverse engineering, decompileren en disassembleren en het bouwen van emulatoren
- archiefvormers maken gebruik van tal van verschillende informatiesystemen, waarvan een aantal op maat van de organisatie zijn gesneden of ad hoc ontwikkeld zijn: archieven dienen over tal van emulatoren te beschikken en kunnen bepaalde kosten niet delen met andere archieven

---

<sup>12</sup> J. ROTHENBERG en T. BIKSON, *Digital preservation*; J. ROTHENBERG, *An experiment in using emulation to preserve digital publications*; J. ROTHENBERG, *Avoiding technological quicksand*; J. ROTHENBERG, *Ensuring the longevity of digital information*.

<sup>13</sup> <http://www.archivebuilders.com/aba010.html>

<sup>14</sup> <http://www.rlg.org/preserv/diginews/diginews5-3.html#feature2>

<sup>15</sup> <http://www.rlg.org/preserv/diginews/diginews5-4.html#feature2>

- emulatie van viewers voor gesloten of niet-gedocumenteerde bestandsformaten op basis van reverse engineering is risicovol, zonet onmogelijk. Emulatie van viewers voor gestandaardiseerde en gedocumenteerde formaten is gemakkelijker en veiliger. Wordt emulatie dan toch voorafgegaan door een migratie naar een gedocumenteerd archiveringsformaat?
- gebruikers werken met oude software en kunnen geen gebruik maken van technologische vernieuwingen
- archieven moeten niet alleen digitale data beheeren, maar ook emulatiehardware en -software en bijhorende documentatie
- de haalbaarheid van bepaalde emulatiepistes zal pas in de toekomst blijken.

De promotoren van emulatie als digitale bewaarstrategie schuiven voornamelijk het behoud van het oorspronkelijk computerbestand met alle originele eigenschappen als belangrijkste argument naar voor. Men beklemtoont in het bijzonder de mogelijkheid tot bewaring van de 'look and feel' en de functionaliteiten, terwijl deze eigenschappen bij migratie veelal gewijzigd worden of verloren gaan. Ze stellen niet de vraag of alle 'originele' eigenschappen wel bijdragen tot de archiefstatus van een digitaal object en bijgevolg wel dienen gearhiveerd te worden, hoe de originele 'look and feel' kan gedefinieerd worden en of het behouden van de oorspronkelijke functionaliteiten wel noodzakelijk is. Ze beschouwen digitale bestanden louter als digitale artefacten waarvan alle eigenschappen behouden dienen te worden. Het is niet toevallig dat de grote emulatievoorstanders in de eerste plaats computerwetenschappers zijn. Men mag niet uit het oog verliezen dat archieven andere taken en doelstellingen hebben dan musea en dat archiefwaardering en contextualisering een essentiële taak van archivariissen is.

Niettegenstaande dit alles, blijft emulatie een potentiële strategie die zijn nut voor de archivering van digitale bestanden kan hebben. In ieder geval heeft men nog maar beperkte ervaringen met emulatie als digitale bewaarstrategie. Er zijn bovendien nog maar weinig praktische en grootschalige emulatietoepassingen voor digitale archivering operationeel.

#### 5.2.4. CONVERSIE

Bij conversie worden digitale data overgezet van een lagere versie naar een hogere versie van hetzelfde bestandsformaat. Een voorbeeld is de conversie van een document van MS Word97 naar MS Word2000.

Voordelen:

- data blijven uitvoerbaar en functioneel.

Nadelen:

- data moeten met een hoge frequentie geconverteerd worden (bijv. MS Word 6.0 -> MS Word97 -> MS Word2000 -> MS Word2002 -> MS Word2003)
- eigenschappen worden gewijzigd of gaan verloren waardoor de authenticiteit moeilijker kan worden gegarandeerd
- digitale data blijven in veel gevallen in een producent-, software- of versiegebonden formaat bewaard: men heeft geen enkele garantie over de lange termijnondersteuning van producent- of softwaregebonden formaten.

Conversie is geen praktische lange termijnbewaarstrategie voor digitale bestanden. Conversie wordt bijgevolg zoveel mogelijk vermeden, tenzij er geen andere mogelijkheden zijn.

Bijvoorbeeld wanneer er geen geschikt archiveringsformaat bestaat of er verlies dreigt van essentiële componenten van het archiefdocument.

#### 5.2.5. MIGRATIE

Migratie is de bewaarstrategie waarbij digitale bestanden naar een geschikt archiveringsformaat worden omgezet. Dit wordt momenteel het meest toegepast bij de archivering.

Aangezien geschikte archiveringsformaten bij voorkeur standaardformaten zijn, is dit de bewaarstrategie waarbij de digitale documenten naar een standaardformaat worden omgezet. Standaarden zijn in principe gedocumenteerd, stabiel en niet afhankelijk van één producent. Migratie wordt soms ook wel aangeduid met 'transformatie' of 'normalisatie' wanneer standaarden als doelformaat worden gebruikt.

Voordelen van migratie als bewaarstrategie:

- digitale data worden niet in een producent-, software- of versiegebonden formaat bewaard
- de specificatie van het bestandsformaat is beschikbaar: op basis van deze documentatie kan ten allen tijd een nieuwe viewer geprogrammeerd worden
- beschikbaarheid van omzettingstools: er zijn niet alleen veel omzettingstools op de markt, migraties zijn ook gemakkelijk te realiseren met behulp van wijdverspreide computerprogramma's.

Nadelen:

- deze bewaarstrategie is sterk gebonden aan standaarden. Standaarden hebben echter een aantal nadelen:
  - hun ontwikkelingsproces neemt veel tijd in beslag: standaarden kunnen de marktevolutie niet even snel volgen
  - standaarden worden niet altijd nauwgezet toegepast of geïmplementeerd: standaarden worden soms uitgebreid om extra functionaliteiten mogelijk te maken waardoor de documenten niet meer ten volle uitwisselbaar zijn
  - standaarden ondersteunen nagenoeg geen applicatie-eigen functionaliteiten
  - niet alle standaarden zijn wijdverspreid of hebben voldoende marktpenetratie
  - standaarden hebben geen onbeperkte levensduur
- voor bepaalde bestandsformaten zijn geen geschikte archiveringsformaten beschikbaar
- de oorspronkelijke eigenschappen of functionaliteiten van het bronformaat kunnen maar zelden integraal worden overgezet naar het doelformaat: migratie gaat in veel gevallen met verlies gepaard
- bij elk omzettingsmoment is de authenticiteit van de data bedreigd.

Migratie is momenteel de meest toegepaste bewaarstrategie voor digitale archiefdocumenten. Bij migratie dient men er wel over te waken dat er geen essentiële informatie verloren gaat of de authenticiteit van de digitale bestanden wordt geschonden. Dit is op zich geen belemmering voor de toepassing van de migratiestrategie. Mits een grondige analyse van bron- en doelformaat kan men risico's vermijden en het verlies tot een minimum beperken. Gekoppeld



aan archiefwaardering moet dit tot overname van alle essentiële en zoveel mogelijk incidentele eigenschappen leiden.

Rekening houdende met de grote hoeveelheid digitale data dient een migratiestap geautomatiseerd uitgevoerd te worden. Manuele omzettingen zijn arbeidsintensief en niet altijd even accuraat. Geautomatiseerde migraties stellen een aantal bijzondere eisen aan het omzettingsproces:

- uitgebreide testfase van de procedure en omzettingsoperatie alvorens tot effectieve toepassing over te gaan
- omzettingstool:
  - uitgebreid testen en controleren zodat men zeker is dat in alle scenario's de omzetting correct wordt toegepast
  - voorzien van een foutenopsporingsmechanisme en error-handling
  - kwaliteitscontrole en validatie van de omgezette bestanden
  - registreren van de documenten die niet correct kunnen omgezet worden zodat deze achteraf manueel gemigreerd kunnen worden.

## Besluit

Uit de evaluatie van de mogelijke digitale bewaarstrategieën blijkt dat er geen enkele bewaarstrategie risicovrij is.

De zoektocht naar een passende bewaarstrategie werd jarenlang toegespitst op de vraag of emulatie van de originele software-omgeving dan wel migratie van de digitale archiefdocumenten de beste oplossing was. Beide oplossingen hebben met elkaar gemeen dat ze een bitstream vertalen naar een leesbaar document. Migratie en emulatie doen dit wel op een verschillend tijdstip. Bij migratie gebeurt dit in het heden, terwijl emulatie deze actie naar de toekomst verschuift. Migratie biedt een oplossing aan documentenzijde, terwijl emulatie voor het leesbaarheidsprobleem aan de hard- en/of softwarezijde een oplossing zoekt.

Inmiddels is het inzicht gegroeid dat beide benaderingen elkaar niet uitsluiten. Beide oplossingen zijn complementair in de levenscyclus van een digitaal bestand of zijn meer geschikt voor een welbepaald type digitaal bestand. In het algemeen is emulatie meer geschikt wanneer de 'look and feel' en het gedrag van belang zijn, terwijl migratie volstaat wanneer inhoud en structuur de essentiële componenten van een bestand zijn. Voor een succesvolle emulatie moet de specificatie van de technologie beschikbaar zijn. Ondertussen bestaan ook tal van tussenoplossingen die elementen van de migratie- en emulatieoplossing combineren.

De internationale inzichten wijzen allemaal in de richting van de "Migration on request"-strategie van het CAMiLEON-project en de benadering van de Nationale Archiefdienst van Australië.<sup>16</sup>

---

<sup>16</sup> P. MELLOR, P. WHEATLEY en D. SERGEANT, *Migration on Request - a practical technique for preservation*, <http://www.si.umich.edu/CAMiLEON/reports/mor/index.html>;  
H. HESLOP, S. DAVIS en A. WILSON, *National Archives Green Paper: An approach to the preservation of digital records*, Canberra, 2002,  
[http://www.naa.gov.au/recordkeeping/er/digital\\_preservation/summary.html](http://www.naa.gov.au/recordkeeping/er/digital_preservation/summary.html)

## 5.3. ARCHIVERINGSSTANDAARDEN

### 5.3.1. *Bestandsformaten*

Digitale data worden bij voorkeur in een gestandaardiseerd bestandsformaat opgeslagen. Gestandaardiseerde bestandsformaten zijn in de regel:

- open en gedocumenteerd: hun technische specificatie is beschikbaar. Men gaat ervan uit dat viewers gemakkelijk te programmeren zijn wanneer men over de technische specificatie van het formaat beschikt
- stabiel: standaarden kunnen pas gewijzigd worden na het doorlopen van een procedure
- software-onafhankelijk: standaarden worden ondersteund door software van meerdere producenten en open source initiatieven
- producent-onafhankelijk.

#### A. HIËRARCHIE

In de IT-wereld bestaan tal van standaarden. Om het overzicht te behouden en als uitgangspunt in de keuze van een bepaald bestandsformaat kan men een hiërarchische indeling hanteren.

Bovenaan in de hiërarchie staan de officiële standaarden. Deze standaarden zijn vastgelegd door officiële standaardiseringsorganisaties en danken hun officiële status aan de participatie van een (inter-)gouvernementele organisatie. Bekende voorbeelden hiervan zijn *ISO* (International Organisation for Standardisation), *IEC* (International Electrotechnical Commission), *ITU* (International Telecommunications Union). Daarnaast zijn er veel officiële regionale en nationale standaardiseringsorganisaties.

Onder deze groep standaarden situeren zich de defacto standaarden. De defacto standaarden kunnen in drie subgroepen worden onderverdeeld. De specificaties zijn het resultaat van niet-officiële standaardiseringsinitiatieven (bijv. W3C). Hun beheer is niet in handen van één producent maar van een standaardiseringsinstantie. De open formaten zijn net zoals de specificaties publiek gedocumenteerd, maar hun beheer is in handen van één bepaalde producent. Tenslotte zijn er de gesloten formaten. Deze formaten danken hun status van defacto standaard aan hun wijdverspreidheid maar hun technische specificatie wordt niet vrijgegeven en ze zijn afhankelijk van één producent.

Bij het kiezen van een geschikt archiveringsformaat richt men zich bij voorkeur op de officiële standaarden en de specificaties. Enig pragmatisme is hierbij aanbevolen. De hiërarchie is een belangrijke leidraad maar is niet zaligmakend. De status van officiële standaard garandeert niets op zich. Zo kennen bepaalde specificaties een grotere toepassing dan hun officiële equivalenten (cf. Unicode vs. ISO-10646; XML vs. SGML). Naast de mate van standaardisatie gelden immers nog andere criteria voor geschikte archiveringsformaten.

#### B. GESCHIKTE ARCHIVERINGSFORMATEN

Een geschikt archiveringsformaat beantwoordt bij voorkeur aan volgende criteria:

- gestandaardiseerd: gedocumenteerd, stabiel en niet afhankelijk van één producent

- wijdverspreid en voldoende marktpenetratie
- uitwisselbaar: onafhankelijk van bepaalde besturingssystemen, netwerkprotocollen en applicaties
- voorziet een robuust foutopsporing- en verbeteringsmechanisme: fouten in bitopslag zijn herstelbaar
- mogelijkheid tot systematische en geautomatiseerde validatie
- goed gestructureerde opslag van informatie
- opslag zonder informatieverlies: geen *lossy* compressie
- mogelijkheid tot insluiten van (zelfgedefinieerde) metadatavelden
- in staat om de essentiële eigenschappen van het archiefdocument in tijd over te brengen
- bewaren van de authenticiteit van de archiefdocumenten
- autonoom en zelfvoorzienig
- drager en apparaat onafhankelijke opslag mogelijk: het formaat kan op elke duurzame drager worden opgeslagen
- eenvoudig en gebruiksvriendelijk: het formaat is niet heel complex of ondersteunt geen ingewikkelde functionaliteiten.

Deze criteria zijn belangrijk in de keuze van een bepaald bestandsformaat als archiveringsformaat. Ook bij het toepassen van archiveringsstandaarden houdt men best deze kwaliteitsvereisten voor ogen. Standaarden kunnen immers op diverse wijzen worden toegepast. De meeste archiveringsformaten laten de gebruiker toe om een aantal instellingen en parameters te definiëren. Zo kan men diverse soorten TIFF-, XML- en PDF-bestanden produceren maar niet elk TIFF-, XML- of PDF-document is geschikt om op lange termijn te archiveren. Bij afbeeldingen die als TIFF-bestand worden bewaard, kan JPEG-compressie worden toegepast. Er gaan hierbij niet alleen gegevens verloren, maar voor de reconstructie is men afhankelijk van de overeenstemmende decompressie. De kwaliteit van XML-bestanden is afhankelijk van de nesting en semantiek van de XML-tags. PDF-bestanden bestemd voor lange termijnbewaring zijn bij voorkeur getagd of op zijn minst gestructureerd.

De gearchiveerde digitale bestanden zijn best zo autonoom mogelijk. De afhankelijkheden voor reconstructie worden best tot een absoluut minimum beperkt. Het ontbreken van één noodzakelijke schakel in het reconstructieproces kan immers tot verlies van data leiden. Omwille deze reden worden compressie, encryptie, paswoorden of andere beveiligingsinstellingen zoveel mogelijk vermeden.

Voor bepaalde types digitale informatie zijn (nog) geen geschikte archiveringsformaten beschikbaar. Deze digitale bestanden zijn nauw verbonden met de hard- en softwareomgeving waarbinnen ze werden gecreëerd en kunnen nauwelijks of niet daar buiten worden gebruikt. Dit is momenteel het geval voor bepaalde multimedia-objecten. In dit geval is het aanbevolen om een bestandsformaat te zoeken dat aan zoveel mogelijk criteria van een geschikt archiveringsformaat beantwoordt en waarbij afhankelijkheden maximaal worden vermeden.

### **Aandachtspunt: archiveer geen gecomprimeerde digitale documenten!**

- decompressie is een extra reconstructieschakel die botst met het principe om afhankelijkheden zoveel mogelijk te vermijden
- bij *lossy* compressie gaat informatie en kwaliteit verloren. Voor audio-visuele archiefdocumenten wordt het kwaliteitsverlies, de ruis en/of de vervormingen gemakkelijk auditief of visueel waarneembaar wanneer verschillende opeenvolgende compressie-algoritmes worden toegepast

- het verwerken van gecomprimeerde bitstreams is complexer
- gecomprimeerde digitale data zijn kwetsbaarder dan ongecomprimeerde data. Een fout in een gecomprimeerd bestand leidt sneller tot onherstelbaar verlies
- de compressienoodzaak vloeit meestal voort uit technologische beperkingen (verwerking, opslag, transmissie). Deze restricties zullen ten gevolge van de technologische vooruitgang de komende jaren soepeler worden of zelfs helemaal verdwijnen.

Als compressie onvermijdelijk is, opteer dan voor een *lossless* compressiemethode (zonder informatieverlies) en kies een compressie met een open, gedocumenteerd en gestandaardiseerd decompressie-algoritme.

### 3.3. VOORBEELDEN VAN GESCHIKTE ARCHIVERINGSFORMATEN

TYPE DATA	ARCHIVERINGSFORMAAT
Tekst:	ASCII/UNICODE, TIFF, PDF of XML
Afbeeldingen:	
Raster:	TIFF, PNG
Vector:	SVG
Raster en vector:	CGM
Geluid:	WAV
CAD:	DXF
GIS	GML
Video:	MXF

#### Tips en aanbevelingen:

- beperk het aantal bestandsformaten die binnen de organisatie als archiveringsformaat worden gebruikt
- bewaar data indien mogelijk van bij de creatie in een geschikt archiveringsformaat
- bewaar data niet in een gesloten of ongedocumenteerd formaat
- vermijd het gebruik van compressie (bijv. LZW, JPEG, ZIP in een TIFF-bestand; ZIP in een PDF-bestand)
- verpak data niet in gecomprimeerde formaten (.zip; .tar)
- wanneer de originele formaten niet bewaard worden: vernietig de originele computerbestanden pas na de controle van de omzettingen
- ga na of standaarden wel correct worden toegepast en of de gearchiveerde digitale data wel beantwoorden aan de formele definitie van de standaard.

### 5.3.2. *Dragers*

#### A. DUURZAME DRAGERS

Of digitale bestanden in de toekomst nog raadpleegbaar zijn, is in de eerste plaats afhankelijk van de drager waarop ze zijn opgeslagen. De digitale data worden best op een duurzame

drager opgeslagen. De drager moet in staat zijn om voor een lange termijn gegevens te bevatten en mag niet snel fysiek aftakelen.

De levensduur van dragers wordt in de regel onderzocht op basis van tests waarbij het verouderingsproces wordt versneld terwijl het aantal fouten op de drager wordt gemeten. Op basis van die tests voorspelt men vervolgens de levensduur van de drager wanneer deze in goede materiële omstandigheden wordt bewaard. Hierbij wordt rekening gehouden met het foutopsporings- en verbeteringssysteem. Voor elk type opslagmedium bestaat immers een foutopsporings- en verbeteringssysteem. Deze mechanismen kunnen tot op zekere hoogte fouten op de drager corrigeren zodat de digitale bestanden leesbaar blijven. Het aantal verbeterbare fouten heeft echter een bovengrens. Als die grens wordt overschreden, zijn de computerbestanden onleesbaar. De levensduurtests geven een goede indicatie van de verwachte levensduur van de drager maar zijn op zich geen garantie voor de leesbaarheid van de archiefdocumenten op lange termijn.

Een duurzame drager en de goede materiële bewaring zorgen er enkel voor dat de drager nog de gegevens bevat die er ooit werden opgeplaatst. Of de informatie op de drager nog effectief kan ingelezen worden, is afhankelijk van de beschikbare technologie.

## B. LEVENSDUUR VAN DE TECHNOLOGIE

In de toekomst dient men immers over de nodige hard- en software te beschikken om de informatie op een bepaalde drager in het computergeheugen te laden (o.a. apparaten, stuurprogramma's). Deze technologie veroudert snel en heeft doorgaans een kortere levensduur dan de dragers die digitale data bevatten. In die zin is het irrelevant of een CD-R al dan niet een levensduur van 100 jaar heeft. De kans is heel groot dat binnen 10 of 20 jaar de apparatuur en/of programmatuur om CD-R's in te lezen, niet meer voorhanden is. Voor verschillende diskette- en tapeformaten is dit nu al het geval. Dit geldt overigens voor alle types dragers, zowel optische als magnetische. De keuze van een drager wordt bijgevolg mee bepaald door de beschikbare technologie. Overzettingen naar andere dragers zullen zich bijgevolg opdringen van zodra een bepaalde technologie niet meer beschikbaar dreigt te worden. Door zorgvuldig een stabiele drager en duurzame technologie te kiezen kan men de overzettingsfrequentie tot een minimum herleiden.

## C. ALGEMENE AANBEVELINGEN

In de praktijk worden zowel magnetische als optische dragers als opslagmedium voor digitale archiefdocumenten gebruikt. Voor beide type dragers gelden de volgende aanbevelingen:

- spreid het risico: bewaar indien mogelijk de data op verschillende types digitale dragers. Hou de verschillende soorten optische en magnetische dragers van digitale bestanden wel in de hand zodat het aantal ondersteunde systemen beperkt kan blijven.
- kies dragers met een technologie die zijn betrouwbaarheid en bedrijfszekerheid inmiddels bewezen heeft; vermijd de allernieuwste technologieën die op dit vlak nog niets bewezen hebben
- bewaar data op dragers die niet snel degraderen: kies dragers met een lange levensduur en een robuust foutopsporings- en verbeteringssysteem
- zorg ervoor dat de nodige apparatuur en programmatuur beschikbaar is:
  - fysiek formaat: gebruik gestandaardiseerde dragers die met meerdere types apparaten van verschillende producenten kunnen gelezen worden
  - logisch formaat: beschrijf de drager volgens een standaard bestandssysteem

- maak veiligheidskopieën en bewaar die op afzonderlijke en veilige locaties: hoe groter de capaciteit of densiteit van de drager, des te meer veiligheidskopieën nodig zijn
- bewaar de opslagmedia in goede materiële omstandigheden
- voer regelmatig kwaliteitscontroles uit
- zet de digitale data naar een andere drager over, wanneer:
  - het aantal verbeterbare fouten op de drager sterk stijgt
  - de technologie in onbruik dreigt te raken
- controleer bij het refreshen de integriteit van de overgezette bytestreams (bijv. door checksums te vergelijken)
- zorg bij elk type drager voor een rampen- en herstelplan
- plaats de data in een gestandaardiseerd bestandssysteem en een open, gedocumenteerd en ongecomprimeerd bestandsformaat op de drager.

#### D. MAGNETISCHE DRAGERS

##### **Aandachtspunt: wees waakzaam bij het gebruik van harde schijven als medium voor lange termijnbewaring!**

- gebruik een type harde schijf dat zijn duurzaamheid bewezen heeft: test harde schijven uitgebreid alvorens ze in gebruik te nemen
- kies een type harde schijven die een gestandaardiseerde interface, of zelfs meerdere interfaces, ondersteunt
- kies harde schijven die uitgerust zijn met SMART (Self-Monitoring, Analysis and Reporting Technology)
- zorg voor beveiliging tegen dataverlies: houd een mirror en/of pariteitsinformatie bij (bijv. RAID 5) (bijv. RAID 5)
- harde schijven zijn niet duurzaam, want hebben een relatief korte levensduur (slijtage, warmte)
- de mappen en bestanden zijn in een bestandssysteem van een bepaald besturingssysteem opgeslagen: een dubbele opslag in twee verschillende bestandssystemen (bijv. Windows en Unix/Linux) is een extra veiligheid.

##### **Aandachtspunt: gebruik geen back-uptapes voor archiveringsdoeleinden!**

Back-uptapes zijn in de regel gecomprimeerde kopieën van platformafhankelijke computerbestanden. Back-uptapes zijn waardeloos zonder de originele back-upsoftware en het computerbesturingssysteem en de applicatiesoftware waarmee de digitale data werden gemaakt:

- back-upformaten zijn meestal ongedocumenteerde of gesloten formaten eigen aan één bepaalde producent of back-upprogramma
- back-upbestanden zijn doorgaans gecomprimeerd. Voor de decompressie is specifieke software nodig
- niet alle informatie voor de reconstructie van de computerbestanden is noodzakelijk op de drager opgeslagen. Bepaalde essentiële informatie wordt op de back-upcomputer bijgehouden.

## E. OPTISCHE DRAGERS

### **Aandachtspunt: gebruik geen DVD's als lange termijndrager!**

- de standaardisatie van DVD is nog niet voltooid
- beschrijfbare DVD's zijn niet gemakkelijk uitwisselbaar.

## **5.4. METADATA**

In de vakliteratuur worden metadata gedefinieerd als “gegevens over gegevens”.<sup>17</sup> Het is een begrip dat vooral wordt gebezigd in de digitale wereld, maar dat hoeft helemaal niet zo te zijn. Globaal gezien zijn er twee mogelijkheden om metadata toe te kennen aan een object, namelijk enerzijds door inkapseling, bijvoorbeeld het colofon in een boek, of in een lijst of databank, zoals een catalogus van bibliotheekbeschrijvingen.

Waar vroeger de nadruk vooral kwam te liggen op de beschrijvingskenmerken, moet er in de digitale wereld ook aandacht worden besteed aan een aantal andere vormen van metadata; technische en beheersmatige aspecten winnen aan belang in een digitale wereld, waaronder context, integriteit, authenticiteit, documentatie over de dragers en bestandsformaten, ...

### **5.4.1 Basisprincipes en aanbevelingen voor implementatie**

Een goed metadataset moet een relatief eenvoudig, gebruiksvriendelijk en flexibel systeem zijn, zowel geschikt zijn voor beheers- als voor ontsluitingsdoeleinden.

#### Eenvoud en flexibiliteit

Bij deze oefening zullen de afwegingen moeten worden gemaakt tussen een vrij eenvoudig en beperkt metadatasysteem dan op het volledig uitwerken van een uitgebreid systeem met tientallen of zelfs honderden metadatavelden. Het creëren van beschrijvingsfiches is zeer intensief, zeker wanneer het gaat om zeer uitgebreide schema's met tientallen of zelfs honderden metadatavelden

#### Hiërarchie en structuur

Het introduceren van een aantal niveaus zorgt ervoor dat het systeem een zekere flexibiliteit verkrijgt. De databank kan zowel worden uitgebouwd rond een oneindig aantal bestanden op hetzelfde niveau, als rond een sterk doorgedreven hiërarchische structuur. In principe kunnen een onbegrensd aantal tussenniveaus worden voorzien.

#### Unieke identificatiecodes

Elke groep, document en representatie krijgt steeds een specifieke unieke code toegekend.

---

<sup>17</sup> Wanneer we dit vertalen naar een archiefcontext, is dit “informatie over archief (documenten)”. Voor meer informatie over metadata en context: P.J. HORSMAN, F.C.J. KETELAAR en T.H.P.M. THOMASSEN (red.), *Context. Interpretatiekaders in de archivistiek*, Stichting archiefpublicaties, 2000.

#### Wettelijke bepalingen van invloed

Wettelijke bepalingen, zoals bijvoorbeeld de auteursrechten en naburige rechten, de wet op de bescherming van de persoonlijke levenssfeer, enz. Om deze problematiek te op te vangen moet de metadata die hierop betrekking heeft, worden vastgelegd.

#### Technische documentatie, geautoriseerde en gecontroleerde lijsten of invoer

Bij het invoeren van een aanzienlijk deel van de metadatatavelden, moet erop gelet worden dat de gegevens steeds op een logische en uniforme manier worden ingevuld. Tikfouten of kleine variaties in schrijfwijze brengen de goede werking van een databank in gevaar en zorgen ervoor dat de gebruikers niet alle mogelijke zoekresultaten te zien krijgen. Niet alleen vanuit het oogpunt van de ontsluiting, maar vooral vanuit het beheer kan dit van groot belang zijn. Wanneer niet alle bestandsformaten op een coherente en uniforme manier worden beschreven, kan het zijn dat bepaalde representaties niet worden gemigreerd en dat de bewaring op lange termijn in het gedrang komt.

Daarom dient voorzien te worden in de mogelijkheid om zoveel mogelijk gebruik te maken van technische documentatie en geautoriseerde of gecontroleerde lijsten. Deze kunnen 'in huis' worden opgesteld en onderhouden maar er kunnen ook gestandaardiseerde lijsten bestaan onderhouden door gespecialiseerde organisaties.

Ook voor andere metadatatavelden, zoals bijvoorbeeld geografische omschrijvingen, auteurs, trefwoorden, enz. zouden in principe kunnen worden gekozen uit een lijst met gecontroleerde, eventueel zelfs geautoriseerde gegevensinvoer.

Technische documentatie kan in principe deels automatisch gegenereerd worden door deze vast te leggen tijdens het digitaliseren of door deze rechtstreeks op te halen uit de digitale documenten. Dergelijke methodologie verkleint het risico op menselijke fouten, maar vereist wel een nauwgezette opvolging naar de accuratesse en de uniformiteit van de gegevens.

#### Automatisch ophalen van bepaalde metadata

Desgewenst kan een functionaliteit worden ingebouwd om een aantal metadata automatisch te genereren of op te halen uit verschillende digitale bestanden

#### Interoperabiliteit

Om te zorgen dat een metadadatamodel breed inzetbaar is, dienen er mogelijkheden worden voorzien om de data te kunnen overzetten naar een andere databank gebaseerd op andere metadatastandaarden. Dit duidt men aan als "mapping" of interoperabiliteit.

### **5.4.2. Metadata voor datasets in de sociale wetenschappen**

Harmonisering van de data en de metadata moet van bij de aanvang ingebed zijn in een archiefdienst. Daarbij moeten internationale geleverde inspanningen en initiatieven gevalideerd worden.

Eén van de meest gekende, en gebruikte, metadatasets is het "Dublin-Core" (DC) model<sup>18</sup>. Deze wordt erkend als een officiële beschrijvingsstandaard door de *International Organization for Standardization*<sup>19</sup> en specifiek ontwikkeld met het oog op het uitwisselen van informatie op het internet.<sup>20</sup> Het is opgebouwd rond 15 optioneel in te vullen velden, die bovendien nog kunnen worden aangevuld met andere velden afkomstig uit andere metadatasystemen. Het model is flexibel en vrijwel onmiddellijk inzetbaar, wat een snelle implementatie mogelijk maakt.

---

<sup>18</sup> <http://dublincore.org/>

<sup>19</sup> ISO 15836:2003

<sup>20</sup> P. HORSMAN, "Metadata and archival description", 2000. [http://www.euan.org/euan\\_meta.html](http://www.euan.org/euan_meta.html) (11 april 2006)



De relatieve eenvoud en een sterke gerichtheid op het gebruiksgemak bij het opzoeken zorgde ervoor dat deze standaard al gauw brede toepassing kreeg in het erfgoedveld, vooral in bibliotheken en musea. Aan DC zijn er echter ook een paar zwakke punten verbonden, zo worden alle objecten naast elkaar beschreven, zonder enige hiërarchische structuur. De omschrijving van de velden is vaak wat vaag, er is geen voorziening voor de uniformiteit van de gegevensinvoer en bijgevolg kunnen de velden op verschillende manieren worden ingevuld.

PREMIS (gebaseerd op OAIS) hanteert het uitgangsprincipe: één document kan worden weergegeven door middel van verschillende verschijningsvormen of representaties. Het doel van de representaties is om een bruikbare weergave van het document te garanderen op de lange termijn. Aan elk document moet ten allen tijde minstens één representatie zijn gekoppeld.<sup>21</sup>

In het archiefwezen is ISAD-G een internationale standaard voor metadata. Deze is sterk ontwikkeld op het terrein van het ontstaan en de vorming van archiefstukken met alle daarbij behorende bureaucratische processen, maar veel minder op het gebied van het documenteren van bestanden, met name ten aanzien van IT-specificaties. Geen enkele van de bestaande archiefdiensten werkt met deze standaard of metadatamodel.

De data-archieven kennen een internationale standaard voor metadatasystemen, het **Data Documentation Initiative** van de internationale organisatie van data-archieven IASSIST. Sedert najaar 2007 is versie 3.0 beschikbaar.<sup>22</sup>

Dit DDI is geschikt voor verschillende soorten bestanden (database én tekstbestanden, multimediabestanden en websites) en kent lokale varianten. In Nederland werd een Nederlandse DDDI (Dutch DDI) ontwikkeld die zich uitstekend leent voor het beschrijven van zowel sociaal-wetenschappelijke als historische databestanden.

Het Data Documentation Initiative is een internationaal initiatief om een standaard te hebben voor alle technische documentatie voor de beschrijving van sociale data afkomstig uit wetenschappelijk onderzoek. Alle DDI-specificaties zijn geschreven in XML.

De meeste XML-schema's en DTD / Document Type Definition zijn beschikbaar via de DDI-website en kunnen daar worden afgehaald.<sup>23</sup> Enige vertrouwdsheid met XML wordt verondersteld maar alles is goed gedocumenteerd.

Omwille van de vertrouwdsheid van een groot aantal instellingen met Dublin Core, is hiermee rekening gehouden. De meeste velden van de Dublin Core element set zijn gemapped aan de DDI Codebook DTD.<sup>24</sup>

Verder bestaat ook **MetaDater : Meta Data Management and Production System for Social Science Surveys**

Ook hier zijn de doelstellingen

- de ontwikkeling van een generiek datamodel en standaard voor de beschrijving van grootschalige surveys, hoofdzakelijk gericht op landoverschrijdende enquêtes of databevragingen

<sup>21</sup> "Data Dictionary for preservation metadata: Final report of the Premis Working Group", 2005, 1-3.  
<http://www.oclc.org/research/projects/pmwg/premis-final.pdf> (11 april 2006)

<sup>22</sup> Zie: <http://www.icpsr.umich.edu/DDI/index.html>

<sup>23</sup> <http://www.ddialliance.org/dtd/index.html>

<sup>24</sup> <http://www.ddialliance.org/related/dc.html>

- dit ondersteunen door de ontwikkeling van softwaretools om deze metadata vast te leggen en verder te beheren.

Het MetaDater project wordt gesteund door de Europese commissie onder het 5th Framework Programme.<sup>25</sup> Alle grote Europese data-archieven zijn opgenomen in dit consortium.

Metadater ontwikkelde

- een conceptueel data model, compatibel met DDI
- een relationeel data model dat in database systemen zoals Oracle of MySql kan worden ingelezen
- diverse softwaremodules, zodat bijvoorbeeld uitwisseling en conversie naar de XML schema's van DDI mogelijk is.

Algemeen kan men stellen dat er op internationaal vlak voldoende initiatieven bestaan in de ontwikkeling van metadata element sets voor data-archieven waar bij de werking van een Belgisch data-archief goed kan worden aangesloten.

## **Chapitre 6. Modalités institutionnelles de création du service**

---

L'un des points cruciaux au moment d'envisager la création d'un centre d'archivage est l'aspect institutionnel de son fonctionnement. Trois questions se posent : quel serait le statut institutionnel du service d'archivage ? Quel serait son mode de financement ? Et quels seraient les organes essentiels à son bon fonctionnement ?

Au stade de l'étude de faisabilité, il nous est impossible de prédire les choix qui seront faits. Dans ce chapitre, nous offrons donc quelques pistes éclairées par les expériences étrangères que nous avons analysées et les échanges avec les différents acteurs potentiels que nous avons rencontrés.

Il est clair que la dimension fédérale de notre pays et en particulier la répartition des compétences en ce qui concerne la recherche entre des entités fédérées (tant régionales que communautaires) et fédérales devront être prises en considération comme elles l'ont été en Allemagne, même si le paysage institutionnel de ce pays est radicalement différent du nôtre et implique la recherche de solutions propres.

### **6.1. STATUT**

#### **6.1.1. Service dédié d'une administration préexistante**

Une première option serait de créer le service d'archivage au sein d'une administration préexistante. Le service d'archivage ferait partie d'un organe étatique déjà constitué à qui seraient alors confiés une tâche et des moyens supplémentaires.

Cet organe devrait cependant remplir deux conditions. Il faudrait premièrement qu'il s'agisse d'un organe faisant partie de l'Etat fédéral et représentant de toutes les communautés du pays. Nous l'avons vu, l'un des intérêts pressentis comme majeurs d'un service d'archivage est de donner une visibilité internationale à la recherche en sciences sociales en Belgique et d'offrir un interlocuteur national aux institutions scientifiques internationales. La deuxième condition liée à l'organe administratif susceptible d'accueillir le système d'archivage serait qu'il ait déjà

---

<sup>25</sup> <http://www.metadater.org/>

une vocation, des missions ou des activités proches de celle du service, de façon à lui apporter dès le début une expertise adéquate.

Trois administrations sont à première vue susceptibles de remplir ces conditions : Direction générale Statistique et Information économique ou Institut National des Statistiques (INS), les Archives générales du Royaume et Archives de l'État dans les Provinces (en un mot les Archives de l'État) et notre commanditaire : la Politique Scientifique Fédérale.

- L'INS s'apparente au projet de service d'archivage dans le sens où il s'attèle à récolter des informations chiffrées, à les recouper et à les rendre utilisables dans un but statistique. Il fait partie du SPF Economie et a donc une portée nationale. Cependant, de par son organisation et ses missions, il se limite au traitement de données *chiffrées* et uniquement *économiques* pour une utilisation par les *organes publics* (et non les chercheurs individuels). De plus, l'archivage dans la durée ne fait pas partie de ses missions.
- Les Archives de l'Etat sont, quant à elle, un Établissement scientifique fédéral qui fait partie du Service public fédéral de Programmation Politique scientifique. Elles veillent à la bonne conservation des documents d'archive produits et gérés par l'autorité, mais elles ont également comme but essentiel la mise à disposition du public de ces données. Elles ont notamment mis en place un système de consultation de données en ligne, et effectuent de la recherche scientifique au niveau de l'archivistique et de la conservation. Elles semblent donc remplir un rôle très proche de ce que serait le rôle du service d'archivage en sciences sociales.

La création de ce service au sein d'une administration existante présenterait l'avantage d'éviter la création d'un nouvel organe fédéral, avec les négociations et les coûts que cela impliquerait, tout en conservant le service au sein de l'administration nationale. Il pourrait également profiter des données administratives déjà disponibles. Il faudrait alors s'assurer de la présence de suffisamment de ressources scientifiques et de l'apport de spécialistes dans le domaine des sciences sociales. Par ailleurs, il n'est pas évident que les Archives de l'Etat soient les plus idoine à servir ce qui est l'intérêt principal de la plateforme d'archivage des données de recherche en sciences sociales. En effet, la mission de cette plateforme serait essentiellement l'aide à la diffusion des données et, à terme, le développement de la recherche en sciences sociales belge. Un service plus proche des milieux de la recherche, plus en lien avec les préoccupations de ce monde particulier est peut-être à préférer comme l'affirmait le représentant suédois lors de notre séminaire international d'octobre 2007.

- La politique scientifique fédérale pourrait, pour cette dernière raison, être préférée aux Archives de l'Etat, à conditions qu'elle en ait les moyens et les ressources. Sans doute serait-elle alors invitée à se doter d'une expertise en matière de gestion d'une plateforme d'archivage et de mise à disposition des données contenues dans les banques de données ainsi constituées. Des liens étroits devraient également être tissés avec les institutions académiques y compris leurs structures faïtières comme le FNRS et son homologue flamand et la coopération avec les administrations commanditaires de recherches devrait être garantie<sup>26</sup>.

### **6.1.2. Service faisant partie d'une université**

Une autre possibilité serait de créer le service d'archivage au sein d'une université. C'est le cas de l'UKDA, et c'est également le choix qui a été fait en Suède. Le processus impliquerait un

<sup>26</sup> Voir à ce sujet la conclusion du chapitre 7 sur les modalités juridiques.

appel d'offre débouchant sur le choix de l'établissement et un financement complémentaire pour celui-ci. Il est à noter qu'en Belgique, l'initiative de la création d'un service d'archivage en sciences sociales a déjà été prise par l'Université Catholique de Louvain, dont le centre (le BASS - Belgian Archives for the Social Sciences), après une longue période d'inactivité faute de financements, semble à nouveau l'objet de l'attention académique.

L'avantage de créer le service au sein d'une université est l'apport en connaissances et en personnel qualifié que cela implique, y-compris via l'emploi à temps partiel de doctorants ou d'académiques. D'autre part, le centre bénéficierait dès le départ des collections de données détenues et créées par l'université qui l'accueille.

Cependant, lorsque cette possibilité a été évoquée avec le directeur de l'UKDA, M. Schurer, celui-ci l'a déconseillée. Il invoque des difficultés d'ordre psychologique et culturel : une certaine rivalité entre universités pourrait amener à limiter l'échange de données, et un centre lié à une université particulière court le risque d'être accusé de favoriser cette université. Au cas où le choix porterait sur une seule université, le problème en Belgique se poserait de manière accrue et ce sur le plan communautaire, car le choix d'une université et donc d'une langue est susceptible d'influencer la collaboration avec les universités de l'autre communauté. De plus, M. Schurer minimise les avantages de cette formule ; selon lui, les connaissances et le personnel peuvent être trouvés de façon tout aussi effective via une collaboration avec un ou plusieurs établissements sans pour autant en faire partie. Enfin, étant donné qu'un grand nombre de données intéressantes sont détenues par les administrations publiques, l'intervention d'une université nécessite des efforts particuliers pour créer un lien de confiance avec les administrations de façon à garantir un échange efficace des données.

### **6.1.3. Service indépendant**

Enfin, une troisième possibilité serait la création d'un nouveau service via un accord institutionnel de coopération entre l'Etat fédéral, les Régions et les Communautés, dépendant des choix de partenariats qui seront faits. Cela impliquerait également un financement et un statut autonomes, voire la création par la loi d'une concession de service public. Il est à noter que l'accord de coopération sera en toute hypothèse nécessaire au cas où le choix porterait sur une administration fédérale dans la mesure où la recherche est de la compétence de tous les niveaux de pouvoir en Belgique.

### **6.1.4. Conclusions**

Différentes possibilités existent donc, chacune comportant ses avantages et ses inconvénients. Des formules mixtes de gestion pourraient également être envisagées (partenariat entre les Archives de l'Etat et la Politique Scientifique Fédérale, gestion par un centre public mais décentralisation des activités auprès d'une université,...). Dans tous les cas, le succès du service ne sera assuré que s'il réalise une collaboration efficace avec les organes de collecte de données existants (INS, Archives du Royaume, BELSPO, universités, SPF,...). Pour être efficace, le service devra s'insérer parfaitement dans la politique nationale de l'archivage.

## 6.2. FINANCEMENT

### *De financiële haalbaarheid van digitale archiveringsdiensten*

Publicaties die aandacht besteden aan het kostenaspect van digitaal archiveren, stellen vast dat het (nog) niet mogelijk is een volledig of betrouwbaar overzicht te geven van alle mogelijke kosten, die bij langetermijnbewaring kunnen optreden.

Een van de geconstateerde problemen is dat het praktisch onmogelijk is om de kosten, die nodig zijn voor het bewaren op zichzelf te scheiden van de kosten, die voor het toegankelijk maken van de data nodig zijn.

Daarbij spelen ook de overheadkosten van het instituut waar de digitale archivering wordt opgenomen, in het bijzonder wat betreft de IT-infrastructuur, zowel in materieel als in personeel opzicht. Zelfs wanneer de kosten van digitale archivering beperkt worden tot opslagkosten, blijft het moeilijk deze vast te stellen.

Stephen Chapman publiceerde in 2003 een onderzoek naar de prijsvorming van digitale archivering. Hij constateerde dat in beide gevallen de opslagkosten door een aantal variabelen bepaald wordt:

- het overeengekomen serviceniveau,
- het soort depot en de wensen van de eigenaar met betrekking tot het aantal collecties, het aantal bestanden,
- het aantal versies en de variatie in formaten.

Kevin Ashley (verbonden aan het NDAD) heeft enige jaren geleden in een publicatie met betrekking tot het kostenaspect van het digitaal archiveren aangegeven dat in hoofdlijnen zeventig procent van de kosten uit arbeidskosten bestaan. Dat betreft alle soorten activiteiten.

1. De meeste tijd wordt besteed aan wat hij noemt 'depositor liaison': het contact en overleg met de opdrachtgever/beheerder van de data. Met andere woorden: het boven water krijgen, van de metadata en contextinformatie.
2. De grootste kostenpost daarna wordt gevormd door kapitaal- en onderhoudskosten voor de hard- en software ten behoeve van de ontsluiting.

Zoals door meer experts is vastgesteld: de grote kosten liggen niet zozeer in de opslag op zichzelf. Het volume van de opslag is een relatief inelastische kostenpost.

In 2006 werd in Nederland onderzoek verricht naar de haalbaarheid van een archiefdienst voor digitale academische databestanden.<sup>27</sup>

Dit Nederlandse project was gericht op het achteraf of retrospectief archiveren van data. Ook zij hebben in hun conclusies geen algemeen overzicht van de kosten van de langetermijnbewaring van onderzoeksdata kunnen geven.

De uitvoering van de werkzaamheden van alleen maar de inventarisatie, de selectie en de classificering op data-clusterniveau, gebaseerd op 1.500 dataclusters, is berekend op ruim 15.000 euro. Deze berekening komt uit op een inzet van 253,8 uur, ofwel van bijna 32 werkdagen, op het niveau van een data-archivist en een IT-medewerker.

De Nederlandse onderzoekers benadrukken dat dit niet alle werkzaamheden betreft. Daarbij is de volgende inzet niet meegerekend:

- Extern uitgevoerde dienstverlening
- Ontwikkeling van procedures
- Doorlopende tijdskosten: communicatie en Bitfaciliteiten
- Doorlopende beheerskosten

---

<sup>27</sup> Heiko TJALSMA, *Archiveren van Digitaal Academisch Erfgoed: een verslag als voorbeeld*. DANS studies in digital archiving, Den Haag, 2006, 70 p.

Deze studie concludeert dat de dienstverlening van een data-archief altijd als maatwerk zal worden aangeboden. Er is keuze uit een aantal modules mogelijk, maar vooral ook de mate waarin een opdrachtgevende instelling meer of minder zelf wil doen tijdens het project kan variëren.

Alleen wanneer een Belgisch data-archief een volledig overzicht kan geven van de hoeveelheid te verrichten werk, is het mogelijk een verantwoorde schatting te doen van de kosten.

De vraag naar de financiële haalbaarheid van digitale archiveringsdiensten is niet met een exacte cijfers of positief / negatief te beantwoorden.

## **6.3. ARCHITECTURE**

Au niveau interne, les expériences étrangères nous enseignent qu'un certain nombre d'organes seraient nécessaires pour réaliser efficacement la tâche qui serait assignée au centre d'archivage.

### ***6.3.1. Service d'acquisition et d'évaluation de la qualité***

Un service du centre devra prendre en charge l'acquisition et l'évaluation des données acquises. Le succès de l'acquisition sera une des données cruciales au succès du centre dans son entièreté. Il est nécessaire d'entreprendre une démarche active de collecte des données, via la recherche des données disponibles, des négociations avec les possesseurs de ces données, la création d'un climat de confiance et d'habitudes d'échange pour les données institutionnelles, etc. Si le choix est fait d'imposer le dépôt des données aux chercheurs subventionnés, il est nécessaire de s'assurer que l'obligation soit bien respectée. Un échange de données avec les centres étrangers doit également être assuré.

Cependant, il peut aussi être nécessaire d'assurer une évaluation de la qualité des données obtenues. A ce niveau-là, deux choix sont possibles : recueillir toutes les données offertes ou faire une sélection. Le premier choix n'est pratiquement viable qu'à la condition d'en avoir les moyens techniques et humains, ou d'offrir un service qui permette le dépôt de données par le chercheur lui-même, comme les systèmes de dépôt en ligne que mettent en place les Pays-Bas et le Royaume-Uni. Dans le cas contraire, il est nécessaire d'opérer une sélection parmi les données offertes pour dépôt, sélection basée sur des critères à définir, tels que la pertinence des données par rapport au but du centre, leur potentiel de réutilisation, ou leur qualité scientifique. Dans ce cadre, il est évident qu'une des tâches devra être de contrôler le respect de la législation de protection des données à caractère personnel et de celle du droit d'auteur.

Enfin, le service d'acquisition et d'évaluation des données pourrait également se charger de la transformation des données collectées (anonymisation et uniformisation des données, ...), dans le but de les rendre directement accessibles aux utilisateurs.

### ***6.3.2. Service de support aux utilisateurs***

Un deuxième service qui, d'après les expériences étrangères, semble avoir son importance, est celui qui serait dédié au support des utilisateurs du service d'archivage. Cela couvre les utilisateurs en amont aussi bien qu'en aval, les déposants et les ré-utilisateurs. Pour les

déposants, il s'agira des les guider dans la préparation de leurs données pour l'archivage, de façon à minimiser les tâches du service d'acquisition ; l'aide aux ré-utilisateurs couvre l'information quant aux données disponibles, à leur localisation (y-compris les données qui sont mise à disposition par d'autres services), aux obligations liées à la réutilisation, et éventuellement une aide pour une réutilisation efficace des données fournies. Des services à valeur ajoutée facilitant l'accès aux données, y compris celles présentes dans des bases de données d'institutions similaires en Europe et l'intégration des données provenant de diverses recherches devront être mis en place. Un système de helpdesk, un moteur de recherche spécifique, ou encore à un système de liens vers des catalogues documentaires spécifiques pourraient apporter cette valeur ajoutée.

### **6.3.3. Comité de gestion**

Il semble évident qu'un comité de gestion devra prendre en main le fonctionnement global du service. Ses tâches comprendraient notamment le fonctionnement général, les relations avec les autorités de financement, le recrutement, l'évaluation et la définition des orientations à prendre par le centre. Ce point a été souligné lors de notre séminaire de réflexion : pour que le centre puisse se construire et évoluer efficacement dans un monde de la recherche en mouvement, il est nécessaire de se poser sans cesse la question de la justification de son existence et de remettre régulièrement ses buts en question.

Cet organe pourrait être composé des directeurs des différents départements du service d'archivage ainsi que des représentants d'une éventuelle autorité de tutelle. On s'interroge sur l'intérêt d'y faire entrer des représentants des Régions et Communautés ou des institutions représentantes de ces Régions et Communautés.

### **6.3.4. Comité de supervision**

Enfin, un dernier organe de supervision pourrait être créé. Il réunirait, outre les représentants du centre, toutes les parties ayant un intérêt dans le service rendu : des représentants des utilisateurs et des fournisseurs de données, de l'autorité de financement, d'autres centres d'archivage en lien avec le service, etc. Ce comité se réunirait ponctuellement dans le but d'offrir un aperçu global de l'environnement dans lequel le service évolue et des attentes de ses partenaires. Grâce à un tel apport, la définition des objectifs à atteindre par le comité de gestion pourrait se faire en toute connaissance de cause.

## **Chapitre 7. Modalités juridiques du dépôt et de la réutilisation des données issues de la recherche**

---

Le dépôt et la réutilisation de données issues de la recherche posent des problèmes juridiques à différents niveaux. Ils concernent la protection des données à caractère personnel, les droits d'auteur et la législation sur l'archivage des données de l'Etat.

### **7.1. PROTECTION DES DONNÉES À CARACTÈRE PERSONNEL**

### 7.1.1. Cadre juridique

La loi belge sur la protection de la vie privée à l'égard des traitements de données à caractère personnel<sup>28</sup> constitue le cadre juridique de protection de la vie privée dans notre pays. Elle définit une série de règles limitant l'accès et le traitement des données dites « à caractère personnel », c'est-à-dire susceptibles d'être, directement ou indirectement, identifiées à une personne physique. Elle a été modifiée le 11 décembre 1998<sup>29</sup> pour rencontrer les exigences européennes en la matière<sup>30</sup>, et se complète d'un arrêté royal d'exécution datant du 13 février 2001<sup>31</sup>.

La loi s'applique à tout traitement automatisé de données personnelles. Elle instaure une série d'obligations lors de la collecte de données à caractère personnel: conditions restrictives pour l'acquisition de ces données, information de la personne concernée, limitation du traitement à ce qui est nécessaire pour réaliser le but recherché, déclaration à la Commission de protection de la vie privée, etc.

En particulier, la loi stipule en son article 4 que des données à caractère personnel doivent toujours être *collectées pour des finalités déterminées, explicites et légitimes, et ne pas être traitées ultérieurement de manière incompatible avec ces finalités, compte tenu de tous les facteurs pertinents, notamment des prévisions raisonnables de l'intéressé et des dispositions légales et réglementaires applicables*. Selon l'interprétation qui est faite de cette disposition dans l'exposé des motifs<sup>32</sup>, la loi laisse donc la possibilité au responsable de traitement de réutiliser les données qu'il détient soit pour réaliser une finalité qui soit compatible avec le but de la collecte, soit en modifiant fondamentalement la finalité, mais dans ce cas il s'agira d'un nouveau traitement au sens de la loi. Ce principe de compatibilité traduit l'exigence de transparence des traitements : chaque personne doit pouvoir prévoir l'utilisation des données la concernant.

La loi sur la protection de la vie privée accorde un régime de faveur aux traitements de données qui sont destinés à effectuer des opérations statistiques, historiques ou scientifiques : d'une part, leur traitement est réputé compatible avec les finalités pour lesquelles les données ont été collectées, et d'autre part, les données peuvent être conservées au-delà de la durée nécessaire à la réalisation des finalités d'origine. En pratique, cette déclaration de compatibilité des finalités implique qu'il ne faudra pas, comme lors de la création d'un nouveau traitement, justifier d'une base légale pour la collecte, informer les personnes collectées de ces nouvelles opérations ou effectuer une nouvelle déclaration à la Commission de protection de la vie privée.

Ces deux exceptions ne seront cependant accordées que si le traitement ultérieur respecte les conditions énumérées dans l'arrêté royal du 13 février 2001<sup>33</sup>. Cet arrêté énonce le principe selon lequel les recherches statistiques ou scientifiques devront s'effectuer autant que possible sur base de données anonymes, c'est-à-dire des « données qui ne peuvent être mises en

---

<sup>28</sup> Loi du 8 décembre 1992 relative à la protection de la vie privée à l'égard des traitements de données à caractère personnel, *M.B.* 18 mars 1993

<sup>29</sup> *M.B.* 3 février 1999

<sup>30</sup> C'est à dire la Directive 95/46/CE du 24 octobre 1995 du Parlement européen et du Conseil relative à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, *J.O.C.E.* n° L281 du 23 novembre 1995, pp. 0031-0050.

<sup>31</sup> Arrêté royal du 13 février 2001 portant exécution de la loi du 8 décembre 1992 relative à la protection de la vie privée à l'égard des traitements de données à caractère personnel, *M.B.* 13 mars 2001

<sup>32</sup> Voir à ce propos l'analyse de T. LEONARD, et Y. POULLET, « La protection des données à caractère personnel en pleine (r)évolution », *J.T.*, 1999, p. 377, n°30

<sup>33</sup> Pour une analyse détaillée des exigences de l'arrêté royal, voir C. De Terwangne et S., Louveaux, « Protection de la vie privée face au traitement de données à caractère personnel : le nouvel arrêté royal », *J.T.*, 2001, p.457



relation avec une personne identifiée ou identifiable »<sup>34</sup>. Ces données n'étant plus des données à caractère personnel, l'exception s'applique sans autre condition.

S'il n'est pas possible d'atteindre le ou les buts recherchés avec des données anonymes, l'arrêté royal impose, autant que possible, l'usage de données codées, c'est-à-dire des données « qui ne peuvent être mises en relation avec une personne identifiée ou identifiable que par l'intermédiaire d'un code »<sup>35</sup>. Ce sera par exemple le cas lorsqu'il sera nécessaire de mettre en relation les données collectées à propos d'une même personne. Une série d'obligations sont alors prévues pour tenter de garantir un codage efficace des données en fonction des hypothèses : traitement ultérieur par le responsable du traitement initial, par un tiers, ou par un sous-traitant.

Si la réutilisation des données sous forme codée ne permet pas d'atteindre les buts historiques, statistiques ou scientifiques recherchés, le traitement de données à caractère personnel sous forme non codée est admis sous une série de conditions plus strictes que celles auxquelles est soumise une collecte primaire dans le même but. Cette rigueur est destinée à compenser la perte de maîtrise des données par la personne concernée et l'avantage du responsable de traitement qui n'a plus à collecter et contrôler les données. Le régime de l'arrêté royal impose une information stricte de la personne concernée et son autorisation expresse à la réutilisation des données. Deux exemptions sont prévues pour les cas d'utilisation de données publiques ou lorsque cette procédure se révèle impossible ou requiert des efforts disproportionnés.

### **7.1.2. Problème**

Les enquêtes en sciences sociales concernent, par définition, des personnes. Si dans de nombreux cas, ces personnes restent anonymes, il arrive qu'elles soient identifiées, soit directement, soit indirectement. Des enquêtes anonymes mais au contenu trop précis peuvent amener à une identification des répondants par recoupement avec des informations publiques: une date de naissance accompagnée d'un lieu de résidence est suffisante et même si le nom n'est pas cité dans l'enquête, toutes les informations se rapportant à la personne deviennent alors des données à caractère personnel.

Dans cette situation, le consentement obtenu par le chercheur initial lors de sa collecte des données auprès de la personne concernée ne couvre pas la réutilisation des données. Cette réutilisation sera donc illégale dès lors que les données sont identifiables. Si elles sont codées, des conditions strictes et lourdes d'un point de vue organisationnel devront être respectées (codage via un tiers,...).

D'autre part, le problème ne concerne pas uniquement le chercheur souhaitant réutiliser les données; la collecte des données par un service d'archivage, comme celui dont notre projet étudie les conditions de faisabilité, constitue déjà un traitement au sens de la loi, et le service sera responsable si le traitement s'est fait de façon illégale.

### **7.1.3. Solutions proposées**

#### **A. UTILISATION DE DONNEES ANONYMISEES**

La solution la plus sûre est celle utilisée par l'UKDA: ne traiter que des données anonymisées, c'est à dire impossibles à lier à une personne particulière. De ce fait, les données sont

---

<sup>34</sup> Article 1<sup>er</sup>, 5° de l'arrêté royal.

<sup>35</sup> Article 1<sup>er</sup>, 3°.

anonymisées par le chercheur lui-même avant leur transmission au centre, ou par le centre lorsque cela n'a pas été le cas.

L'UKDA s'entoure en outre de précautions supplémentaires:

- Il impose au déposant de garantir, dans le contrat de dépôt, la légalité de la collecte originale, y compris en ce qui concerne la protection de la vie privée des répondants. Cela suppose, à notre avis, que des vade-mecum soient élaborés par l'organisme en charge du dépôt et ce lien avec la Commission de Protection de la Vie Privée (CPVP) pour permettre de comprendre la portée des dispositions de la loi de protection des données voire qu'un système de labellisation des centres de recherche soit mis sur pied qui permettent la vérification a priori du respect de ces dispositions et puisse attester de leur respect);
- lors de la mise à disposition des données, le ré-utilisateur s'engage à ne pas chercher à identifier les données;
- Il suit la réutilisation des données via un contrôle effectif des enregistrements et des téléchargements, garantissant ainsi un suivi responsable en cas de violation du contrat de dépôt ou de réutilisation, et se protégeant des critiques. On note que ce contrôle constitue un traitement de données à caractère personnel des « ré-utilisateurs » et que ce traitement doit suivre lui-même les principes de transparence de la loi de 1992 et ne servir qu'à la finalité de contrôle;
- Il traite principalement les ensembles de données totalement anonymes et donc certaines variables peuvent être supprimées ou modifiées pour limiter le risque d'atteintes à la confidentialité (par exemple, des variables géographiques). Si des informations plus précises sont nécessaires, des obligations particulières seront attachées à leur fourniture via une licence spéciale (*cfr. supra*). Ces obligations comprennent une justification de la nécessité d'utiliser ces données, la signature du chercheur ou de l'institution qui en a la responsabilité, ainsi que la permission explicite du propriétaire des données préalablement à chaque réutilisation.

En outre, l'UKDA, en tant que responsable de traitement de données à caractère potentiellement personnel (ne serait-ce qu'au moment de leur anonymisation), a entré une déclaration de traitement auprès de l'autorité nationale de protection des données, conformément à la législation européenne.

Il semble que ces précautions n'ont pas mis à mal le développement du centre d'archivage anglais. Elles présentent au contraire l'avantage de faciliter la gestion du centre au point de vue juridique et de gagner la confiance des autorités fournissant les données.

## **B. COMPILATION ET ANONYMISATION PAR LE SERVICE D'ARCHIVAGE**

Une autre solution pourrait être envisagée, dans le but d'augmenter l'efficacité et l'attractivité du centre d'archivage. Celui-ci pourrait avoir la possibilité de récolter des données non anonymes et de les compiler avec d'autres données qui lui ont été confiées de façon à obtenir de nouvelles informations susceptibles d'intéresser les chercheurs, avant d'anonymiser les résultats et de les rendre disponibles. Le service d'archivage pourrait de cette façon devenir lui-même producteur de données nouvelles.

Il s'agirait, par exemple, de recevoir les résultats d'une recherche portant sur le comportement d'achat d'une série de familles identifiées par leur composition, leur localisation, l'âge et la profession de leurs membres. Ce type de données peut être qualifié de personnelles car elles

sont très probablement suffisamment précises pour permettre l'identification de la famille. Le centre serait donc tenu, avant de les rendre disponible, d'élargir les catégories ou de supprimer certaines variantes de façon à les anonymiser. Cependant, il pourrait être intéressant, au préalable, de recouper les informations disponibles avec d'autres recherches pertinentes mises à la disposition du centre, portant par exemple sur les niveaux de revenu en fonction de l'âge, de la localisation et de la profession. Le centre pourrait ainsi créer une nouvelle catégorie de données : les comportements d'achat en fonction du niveau de revenu.

Dans cette optique, le service d'archivage effectue deux manipulations sur les données à caractère personnel : une utilisation à des fins de recherche scientifique, et une opération d'anonymisation.

En ce qui concerne l'utilisation à des fins de recherche scientifique, il s'agit en réalité d'une réutilisation des données collectées pour une autre finalité que la finalité initiale. Les conditions entourant cette réutilisation ont été explicitées plus haut. La légalité de la réutilisation des données se base en fait sur une question : le traitement ultérieur est-il ou non compatible avec la finalité initiale ? Dans le cas contraire, il s'agira d'un nouveau traitement, impliquant une nouvelle information de la personne concernée et l'application de toutes les obligations liées au traitement de données à caractère personnel.

Cette compatibilité des finalités doit, selon la loi, s'analyser au regard des éléments pertinents, comme les prévisions raisonnables de l'intéressé<sup>36</sup>. Il est probable que dans la plupart des cas, la personne qui accepte de répondre à une enquête donne son consentement à l'utilisation de ses données dans un but d'étude scientifique. A la condition que le chercheur à l'origine de la collecte n'ait pas précisé un cadre limitatif à ses recherches<sup>37</sup>, et également à la condition qu'il ne s'agisse pas de données sensibles, médicales ou judiciaires, l'utilisation ultérieure des données par le service d'archivage dans un but d'analyse scientifique complémentaire peut donc probablement être considérée comme compatible avec la finalité initiale. Peut-être faudrait-il, pour plus de certitude, que les chercheurs qui collectent des données avec l'intention de les transmettre dans un deuxième temps au centre d'archivage, en informe la personne intéressée.

Le deuxième traitement est l'opération d'anonymisation. Il s'agit également d'un traitement de données à caractère personnel au sens de la loi. L'arrêté royal du 13 février 2001 prévoit en son chapitre II un système de règles lors de la transmission de données pour une réutilisation de données à des fins statistiques, scientifiques ou historique, système qui implique l'intervention d'une organisation intermédiaire responsable du codage des données. Cependant, ce chapitre ne s'applique que dans le cas où le nouveau traitement à des fins statistiques, scientifiques ou historiques n'est a priori pas compatible avec le but de la collecte initiale. Or, dans le cas du système d'archivage, les données auront été collectées dans un but statistique et la réutilisation n'est en quelque sorte que le prolongement de l'utilisation initiale. Par conséquent, il n'est plus nécessaire de faire appel à une entreprise intermédiaire pour l'anonymisation des données.

En résumé, rien ne s'oppose à ce que le service d'archivage rassemble les données collectées, les recoupe et les anonymise avant de les transmettre aux chercheurs intéressés, à la condition que

---

<sup>36</sup> C'est-à-dire le sujet de l'enquête.

<sup>37</sup> On note cependant que, suivant les articles 4 et 9 de la loi, le chercheur devra préciser de manière explicite les finalités déterminées de sa collecte. Il apparaît donc difficile qu'il ne mentionne pas l'objet précis de sa recherche et qu'une clause affirmant que les données collectées pourront servir à des recherches ultérieures en sciences sociales sera jugée sévèrement par la Commission belge de protection de la vie privée.

- l'information fournie par le chercheur initial aux personnes concernées par son enquête ne soit à ce point restrictive qu'elle ne permette pas la réutilisation ;
- ne soient pas collectées des données sensibles, médicales ou judiciaires.

## **7.2. PROTECTION DES DROITS D'AUTEUR**

### **7.2.1. Cadre juridique**

La législation sur les droits d'auteur en Belgique se concrétise par la loi du 30 juin 1994 relative au droit d'auteur et aux droits voisins<sup>38</sup>, qui intègre notamment les directives européennes en matière de protection des bases de données<sup>39</sup>.

La loi prévoit en la matière que l'auteur d'une œuvre littéraire ou artistique (à savoir, la personne physique qui a créé l'œuvre) a seul le droit de la reproduire ou d'en autoriser la reproduction. Lorsque l'œuvre est le produit d'une collaboration, le droit d'auteur existe au profit de tous les ayants droit.

La loi distingue les droits moraux et les droits patrimoniaux. L'auteur ne peut renoncer globalement à l'ensemble de ses droits moraux sur l'œuvre (comme le droit de divulguer l'œuvre, d'en revendiquer ou refuser la paternité, ou le droit de s'opposer à toute modification), mais peut aliéner ses droits patrimoniaux<sup>40</sup>. Lorsque l'œuvre a été créée en exécution d'un contrat de travail ou d'un statut, les droits patrimoniaux peuvent être cédés à l'employeur.

En ce qui concerne les bases de données, la loi qualifie comme telle « un recueil d'œuvres, de données ou d'autres éléments indépendants, disposés de manière systématique ou méthodique et individuellement accessibles par des moyens électroniques ou d'une autre manière ». La loi prévoit une protection par le droit d'auteur à leur profit lorsqu'elles constituent, par le choix ou la disposition des matières, une création intellectuelle propre à leur auteur. Il faut donc faire la preuve d'une certaine originalité dans le choix et la collecte des données. De plus, la loi prévoit une présomption de cession des droits patrimoniaux relatifs aux bases de données créées, dans l'industrie non culturelle, au profit de l'employeur.

La loi prévoit enfin un certain nombre d'exceptions aux droits d'auteur en faveur de l'utilisateur qui poursuit une finalité purement scientifique et non commerciale. Cependant, ces exceptions sont de portée limitée et ne couvrent pas la transmission systématique à des tiers, ou la création d'une nouvelle œuvre sur la base d'une œuvre protégée.

### **7.2.2. Problèmes**

La création d'un centre d'archivage qui mettrait à disposition les données collectées pour une réutilisation pose un certain nombre de problèmes au niveau de l'application des droits d'auteur. Qui possède les droits sur les résultats de la recherche, en particulier dans les universités, qui du chercheur individuel, du laboratoire ou de l'Unité de recherche, de l'Université elle-même détient les droits d'auteur? Qui a le pouvoir de les aliéner, et en quels termes? Les universités devront sans doute veiller à inclure dans leur contrat avec les

---

<sup>38</sup> M.B. 27 juillet 1994

<sup>39</sup> Directive 96/9/CE du Parlement européen et du Conseil, du 11 mars 1996, concernant la protection juridique des bases de données, *J.O.* n°L077 du 27/03/1996, p.00-20 – 00-28

<sup>40</sup> C'est-à-dire ses droits évaluable en argent.

chercheurs et de manière explicite des clauses de cession de droit, sauf les droits non patrimoniaux. Reste une deuxième question, la première résolue : quel sera le statut d'une œuvre secondaire ? A cet égard devrait être prônée l'utilisation de licences « open content », qui facilitent sous réserve de l'obligation de citer les auteurs, la réutilisation des contenus comme nous le préconisons dans les solutions que nous décrivons maintenant.

### **7.2.3. Solutions proposées**

#### **A. IDENTIFICATION DES TITULAIRES DES DROITS D'AUTEUR**

La première question à envisager est celle de la propriété des droits sur l'œuvre. La loi considère la personne qui a créé l'œuvre comme son auteur. Dans le cas des données en sciences sociales, il s'agit du chercheur réalisant l'enquête. Il peut cependant avoir participé à une œuvre collaborative ; chaque auteur a alors un droit équivalent sur l'œuvre, et sa mise à disposition implique un accord de tous.

Lors de la création d'une base de données, le chercheur possède un droit non pas du fait des données elles-mêmes mais sur la partie d'originalité qu'il leur apporte : le questionnaire, les critères de classement, l'analyse des données,...

Il arrive fréquemment que le chercheur ait cédé ses droits patrimoniaux sur l'œuvre dans le contrat de financement. Il peut également avoir garanti à l'organisme de financement une exclusivité sur l'œuvre produite<sup>41</sup>.

Aucune règle générale ne peut être retenue, les contrats de financement ainsi que les acteurs de la recherche variant au cas par cas. Un système d'archivage récoltant des données ne peut se permettre de vérifier les titulaires des droits d'auteur pour chacun des ensembles de données qu'il collecte. Le contrat de dépôt devra donc contenir une clause mettant cette recherche à la charge du déposant. Ce dernier devra garantir d'une part qu'il est le seul titulaire des droits d'auteurs ou qu'il a la permission des autres auteurs et/ou de ses partenaires de recherches de rendre les résultats de la recherche publics, d'autre part qu'il n'enfreint aucune règle d'exclusivité inscrite dans son contrat de recherche ou dans son contrat de financement, ceci pour éviter les recours par la suite.

Une telle disposition ne dispense cependant pas le centre d'archivage d'exercer un contrôle raisonnable de l'application des dispositions juridiques, dans les limites de ses moyens.

#### **B. MISE A DISPOSITION DE L'OEUVRE**

Une fois les titulaires des droits sur l'œuvre identifié(s), un contrat devra régler l'autorisation de mise à disposition de l'œuvre pour consultation et réutilisation. La loi sur les droits d'auteur prévoit que « l'auteur d'une œuvre littéraire ou artistique a seul le droit de la communiquer au public par un procédé quelconque, y compris par la mise à disposition du public de manière que chacun puisse y avoir accès de l'endroit et au moment qu'il choisit individuellement ». Il convient donc de régler par écrit (le contrat de dépôt) la mise à disposition des données. L'auteur ne devra pas céder ses droits sur l'œuvre, mais uniquement autoriser son utilisation dans un cadre précis. Différentes modalités sont possibles quant au type d'utilisation qui sera accordé, mais dans tous les cas les limites imposées devront être précisément définies en sachant que « les dispositions contractuelles relatives au droit d'auteur et à ses modes

---

<sup>41</sup> Voir à ce sujet l'exemple du Règlement général en matière de propriété, de protection et de valorisation des résultats des recherches réalisées au sein des FUNDP, disponible en annexe 9.

d'exploitation sont de stricte interprétation »<sup>42</sup>. Le contrat devra également impérativement être conclu par écrit.

### C. STATUT DE L'ŒUVRE DERIVEE

Le chercheur qui réutilise les données rendues disponibles via une plateforme d'échange apporte un contenu original nouveau : nouvelle analyse, recoupements avec d'autres données,... Son droit d'auteur s'applique à ce contenu. Les droits de l'auteur original ne lui sont pas transférés, il doit donc les respecter : d'une part il doit citer scrupuleusement les références de l'œuvre originale dans son propre travail, et d'autre part il ne peut céder les droits relatifs à l'œuvre originale, y compris en les rendant à nouveau publics (notamment sur une autre plateforme) sans le consentement du titulaire des droits d'auteur. Ces obligations devront être rappelées avant toute réutilisation, et le centre d'archivage doit ici aussi, pour se prémunir contre toute critique, en vérifier l'application dans une mesure raisonnable.

### D. CONCLUSION

La solution la plus sûre pour éviter les conflits liés aux droits d'auteur sur les données confiées est d'imposer aux chercheurs, comme le font la plupart des centres d'archivage, deux contrats écrits :

- un contrat de dépôt avec le fournisseur des données, par lequel celui-ci s'engage à disposer de l'autorisation de tous les auteurs (c'est-à-dire les scientifiques à l'origine de l'œuvre intellectuelle) à mettre ces données à la disposition de la communauté scientifique. Ce contrat devra préciser les limites de cette mise à disposition, qui n'est pas une cession de droits. Sans doute, sera-t-il utile de proposer à ou aux auteur(s), un contrat de type licence ouverte ou « open content » qui garantit le droit de réutilisation des sources par les tiers mais oblige ceux-ci à mentionner l'auteur et à eux-mêmes accepter le système de l'open content vis-à-vis des tiers qui souhaitent avoir accès à l'œuvre dérivée. On peut même imaginer que le ré-utilisateur soit tenu de la même manière de mettre à disposition du service d'archivage l'œuvre dans les mêmes conditions que le ou les auteur(s) originaires.
- un contrat de réutilisation avec celui qui souhaite accéder aux données, par lequel celui-ci s'engage à respecter les droits d'auteur sur l'œuvre originale, via notamment une obligation de citation.

Une série de contrats de ces deux types se trouve à l'annexe 7. Il s'agit des contrats utilisés par les centres d'archivage examinés : le DANS, l'UKDA et le ZA. Ils fournissent de bons exemples qui pourraient servir de base à l'élaboration de contrats spécifiques au centre d'archivage belge, en fonction des choix stratégiques qui seront réalisés (types de données accessibles, types d'accès, types d'utilisateurs autorisés, etc).

---

<sup>42</sup> Article 3 §1<sup>er</sup> de la loi.

## 7.3. OBLIGATIONS LIÉES À L'ARCHIVAGE

### 7.3.1. Cadre juridique<sup>43</sup>

En ce qui concerne les données récoltées par les services publics, la loi sur les archives du 24 juin 1955<sup>44</sup> s'applique. Elle est complétée par un arrêté royal datant du 12 décembre 1957<sup>45</sup>.

La loi sur les archives s'applique aux producteurs d'archives publiques, de tous niveaux institutionnels. Elle leur impose (à l'exception des communes et des organismes d'utilité publique) de déposer aux Archives nationales tous les documents de plus de cent ans, et, à condition qu'ils n'aient plus d'utilité administrative, leurs documents de moins de cent ans. Les pièces ainsi déposées sont publiques, c'est-à-dire librement consultables. De plus, la loi confère à l'Archiviste général du royaume (à savoir, le directeur général des Archives nationales) le pouvoir de contrôler les conditions de conservation des données publiques chez les producteurs.

Ce cadre légal est complété par la législation sur la transparence administrative<sup>46</sup>, consacrant la publicité des documents administratifs. Elle contraint les administrations fédérales à faire une publicité active sur leur organisation, leurs missions et l'exercice de celles-ci. Au terme de cette loi, les autorités ont l'obligation de répondre de manière positive aux demandes de renseignement des citoyens, y-compris aux demandes de consultation et de copie de documents de portée générale, sauf s'ils font l'objet d'une obligation de secret imposée par la loi. Chaque citoyen a donc le droit de consulter et d'obtenir copie des informations qui le concernent personnellement et des dossiers de nature générale qui ne sont pas classifiés.

### 7.3.2. Problème

Un service d'archivage en sciences sociales trouvera un grand intérêt dans l'acquisition de données d'origine publique : résultat d'enquêtes menées à grande échelle, données récoltées par des autorités dans un but statistique ou d'aide à la décision, *etc.* Ces données seraient précieuses pour une réutilisation scientifique. Cependant, elles sont déjà soumises à une obligation d'archivage auprès des Archives nationales, et de fourniture auprès des citoyens. Comment concilier la réalisation de ces obligations avec la création d'un centre d'archivage ?

### 7.3.3. Solution proposée

Il ne semble pas y avoir de conflit entre la réalisation de ces différentes dispositions, à la condition de réaliser les activités d'archivage en collaboration avec le directeur général des Archives nationales, qui conserve son pouvoir de contrôle en vertu de la loi. Une possibilité à envisager serait de créer le centre d'archivage en sciences sociales sous la direction des Archives nationales, qui possèdent déjà une expérience dans le domaine de la conservation et la mise à disposition de données. Il deviendrait ainsi inutile d'archiver les mêmes données à la fois dans le centre créé et aux Archives nationales. Une autre solution est la décentralisation des opérations d'archivage dans le cadre d'un contrat passé entre, d'une part, le service

<sup>43</sup> Lire à ce sujet: R. DEPOORTERE, « Les archives en Belgique : une réalité éclatée », *Archives*, Vol. 34, Numéros 1 et 2, 2002-2003, p.101, disponible à l'adresse

[http://www.archivistes.qc.ca/revuearchives/vol34\\_1-2/34-1&2-Depoortere.pdf](http://www.archivistes.qc.ca/revuearchives/vol34_1-2/34-1&2-Depoortere.pdf)

<sup>44</sup> M.B. 12 août 1955

<sup>45</sup> M.B. 20 décembre 1957

<sup>46</sup> Loi du 11 avril 1994 sur la transparence administrative, M.B. 30 juin 1994

d'archivage visé par notre étude et qui pourrait être placé sous la responsabilité de BELSPO et, d'autre part, les Archives nationales qui pourraient par un hyperlien renvoyer au premier et assureraient la conformité des opérations de ce service aux conditions de la loi sur les archives.

## Technisch besluit

---

Al sinds de jaren 1960 bestaan er archiefdiensten voor sociaalwetenschappelijke data, zowel in Europa en Noord-Amerika. Later zijn daar historische data-archieven en tekstarchieven bijgekomen.

Hun belangrijkste taak is het bewaren en toegankelijk houden van bestanden, waarbij mogelijk hergebruik een belangrijk criterium is. De gehanteerde strategie is meestal conversie naar een software-onafhankelijk of gestandaardiseerd opslagformaat als ASCII of XML. Sociaalwetenschappelijke data-archieven converteren meestal naar een portable file format; tekstarchieven maken gebruik van de markup languages SGML of XML en historische data-archieven gebruiken ASCII en tegenwoordig op experimentele basis ook wel XML. Multimediale bestanden en op internet gepubliceerde databases vragen om nieuwe archiveringsoplossingen.

De data-archieven kennen een internationale standaard voor metadatasystemen, het Data Documentation Initiative. Dit DDI is geschikt voor verschillende soorten bestanden (database én tekstbestanden, multimediabestanden en websites).

Op het gebied van digitale archivering gebeurt bij overheidsarchieven wel het nodige, en bestaat er nationaal zowel als internationaal voldoende expertise maar het archiveren van digitale archiefstukken (electronic records) stelt geheel andere eisen dan voor de analoge of traditionele media en dragers. Vooral vraagt het een andere organisatievorm en andere procedures.

De vraag om prioriteiten stelt zich. Naar de haalbaarheid van retrospectieve digitale archivering voor de wetenschappelijke wereld dient uitgebreid onderzoek gedaan te worden.

Toch bevestigt de bevraging van de gebruikersgroepen grotendeels de urgentie van een retrospectieve archivering. Het risico dat onderzoeksbestanden verdwijnen wordt snel groter door het verdwijnen van de kennis over de te archiveren data. Vooral het documenteren en selecteren van data, die hebben toebehoord aan inmiddels verdwenen vakgroepen of onderzoekers kan veel problemen gaan opleveren. Het probleem lijkt eerder op het vlak van verdwijnende kennis te liggen dan op dat van de leesbaarheid van oudere media.

Daarnaast is een actieve vorm van archivering, minder gebruikelijk in de papieren archiefwereld, zo noodzakelijk om langetermijnbewaring van digitale bestanden te bewerkstelligen. De keuze voor het organiseren van een actieve archivering van alle wetenschappelijke data die voortaan worden verzameld kan allicht gemakkelijker worden gemaakt. De factoren die een rol spelen zijn van technische en documentaire aard, maar vooral van organisatorische en beleidsmatige aard.

Een aanpak voor een toekomstige digitale archiveringsdienst waarbij België de huidige achterstand in de organisatie van een '*Social Science Data Service*, moet de huidige lacune aanpakken. Deze lacune is niet alleen nadelig voor de duurzaamheid van het wetenschappelijk onderzoek in België zelf. Het zal op relatief korte termijn België ook parten spelen wanneer blijkt dat niet kan worden ingestapt in een aantal Europese programma's, of dat Europese richtlijnen of aanbevelingen niet kunnen worden omgezet bij gebrek aan een leidende instantie.



Daarbij gelden volgende conclusies:

1. Noodzakelijk is een goede schatting vooraf van de grootte van het project, vooral gebaseerd op het aantal databestanden en de variëteit daarin, inzicht in de data-infrastructuur en de automatiseringsgeschiedenis.
2. Goede communicatielijnen met de onderzoeksinstellingen omwille van de assistentie die vereist is gedurende de gehele archivering, zelf onontbeerlijk is, speciaal ten aanzien van de inhoudelijke kennis van de databestanden en de context waarin deze zijn ontstaan. Zonder de inschakeling van vakspecialisten is inhoudelijke selectie niet mogelijk.
3. Participatie van de onderzoeksinstellingen lijkt noodzakelijk, maar de intensiteit daarvan kan sterk verschillen.
4. De selectie van de te archiveren databestanden gebeurt best in fasen.
5. Een modulaire aanpak is sterk aan te bevelen, vooral door de keuzemogelijkheden die worden geboden aan de opdrachtgevende instantie.
6. De noodzaak aan een (centrale) instelling in België die zich bezighoudt met advisering en mogelijk ook opslag van wetenschappelijke databestanden van groot belang voor de Europese (vb CESSDA) en internationale netwerken. Vermoed wordt dat een federale instelling een groter impact zal hebben.
7. Studies hebben aangetoond dat 70 % van de kosten ligt in de personeelskost. Omwille van de overhead inzake IT-infrastructuur lijkt het aangewezen aan te sluiten bij een instelling waar deze reeds aanwezig is.
8. Een koppeling aan het DISSCO-project verdient nader onderzoek.
9. De activiteiten van BASS moeten worden uitgeklaard nadat deze instelling sedert 10 jaar de rol niet meer kan spelen.
10. Duidelijke afspraken ten aanzien van sociale databestanden die onder het toezicht vallen van overheidsarchieven



## CONCLUSIONS GÉNÉRALES

Dans ce rapport de recherche, nous avons voulu faire le tour de la question de l'archivage en sciences sociales et fournir une série de réflexions utiles à la mise en place d'un tel service approprié au contexte particulier de la Belgique. Si certaines options demandent encore à être tranchées<sup>47</sup>, nous en arrivons cependant aux recommandations suivantes.

### Un service nécessaire...

Au terme de cette étude de faisabilité, nous arrivons à une nécessité centrale **de mettre en œuvre un service d'archivage des données de sciences sociales en Belgique**. Le service dont nous avons testé la faisabilité tout au long de cette étude a en effet été très bien accueilli et largement plébiscité par les nombreux acteurs que nous avons rencontrés, fortement désireux de voir sa concrétisation aboutir.

Pour les acteurs interrogés, la création d'un tel service permettrait une **meilleure visibilité** et partant, un meilleur positionnement de la recherche en sciences sociales au plan européen et plus largement, international.

Ce service devrait également participer à une **amélioration des échanges scientifiques** entre les chercheurs belges et favoriser de ce fait un développement plus concerté des sciences sociales dans notre pays.

Au niveau des institutions qui administrent et décident de la recherche, un tel service peut également aider à **définir des programmes de recherche tenant compte de l'existant**, éviter un certain gaspillage des moyens nécessaires à la réalisation de grandes enquêtes, par exemple, mais aussi à « **gouverner** » en s'appuyant sur des données de référence relative à l'état social de notre pays.

<sup>47</sup> Nous pensons notamment à la période jusqu'à laquelle rechercher et conserver des données ou au budget libéré pour la réalisation de ce projet.

### **Un service proactif ...**

Si un tel service apparaît nécessaire aux yeux de tous, nous soulignons également qu'il n'aura d'utilité que si il est réellement proactif. Par proactif, on entend un service régulièrement mis à jour par de nouvelles données et qui pour cela démarche régulièrement les dépositaires et les utilisateurs et les informe systématiquement des possibilités offertes par le service.

### **Un service qui ne fait pas double emploi...**

La volonté de mettre en œuvre un tel service va de pair avec le souhait de bien le positionner, faisant en sorte que tant au niveau politique que pragmatique il s'intègre facilement dans le paysage des initiatives déjà existantes, tel les portails et les nombreux instituts de statistiques existant au niveau fédéral, régional ou européen (archives générales du Royaume, Statbel, IWEPS, ...). La **complémentarité** entre ce service à créer et les initiatives en place est un facteur clef pour sa réussite.

### **Un service à déployer en paliers ...**

Les expériences étrangères analysées dans ce rapport nous l'apprennent, la mise en œuvre d'un tel service est une entreprise longue et qui peut être assez demandeuses en moyens. Il est important que le service de recherche se fixe des priorités de développement. C'est pourquoi, nous préconisons un déploiement du futur service **par paliers**.

- Palier 1. Le service d'archivage serait alimenté par de la documentation sur des recherches quantitatives et publiques (rapports de recherche, questionnaires, explication de méthodologie...) sur base de politiques incitatives favorisant les dépôts de données par les chercheurs. Ce référentiel des recherches quantitatives menées en sciences sociales serait accessible au seul monde académique et administratif.
- Palier 2. Le service d'archivage serait complété des données d'enquêtes quantitatives (accompagnées d'un guide d'usage) ainsi que de données provenant de recherches plus qualitatives. Ce deuxième palier pourrait également s'accompagner d'un élargissement du public pouvant accéder au service. Un système d'enregistrement sécurisé au portail pourrait également être mis en place pour accéder à des données plus sécurisées et à des matériaux de base (chiffres collectés, extraits d'entretiens, ...) et également pour permettre aux chercheurs, moyennant certaines conditions, de déposer eux-mêmes leurs données (système d'easy data depositing).

### **Un service reposant sur un contrat entre parties...**

La mise en place d'un tel portail n'est pas non plus sans poser un certain nombre de **questions juridiques** sur le dépôt et la réutilisation des données par des tiers. Ces questions touchent au respect des droits intellectuels de l'auteur original et du propriétaire des données, à la vie privée des sujets de recherche, ainsi qu'aux réglementations d'archivages des données publiques. Les problèmes potentiels peuvent cependant trouver leurs solutions dans l'élaboration de contrats liant le centre d'archivage avec les différentes parties, contrats dont le contenu dépendra des choix stratégiques qui seront faits (types de données collectées, d'utilisations consenties, etc).

### **Un service reposant sur des standards internationaux**

L'utilité d'un tel service est liée à la lisibilité et l'accessibilité des données qui y sont conservées. C'est la raison pour laquelle tant au niveau des **méta-données** que des fichiers de données eux-mêmes, le service devrait œuvrer en faveur de l'usage de standards internationaux.

### **Un service supporté par une culture de la coopération...**

Pour terminer, une forte **culture de partage** de données chez les dépositaires, une bonne collaboration entre le service d'archive et les institutions belges de la recherche et une bonne intégration du dispositif au sein de relais internationaux semblent des ingrédients essentiels pour pouvoir garantir le succès de ce service d'archivage.





## **BIJLAGES / ANNEXES**

### **BIJLAGE A. Feasibility study of a Belgian computerised Archive Service for the social sciences (SODA). English description of the project**

---

#### General goal of this research project

This project aims at studying the possible development of a federal computerised information archiving service for the social sciences in Belgium.

Social sciences researchers often need to collect a large number of qualitative as well as quantitative data, which requires a high human and financial investment. However, most of the time, these data bearing a high scientific value sink into oblivion after their first use.

To solve that problem, many countries have already developed a service in charge of storing data collected and keeping them available for reuse. However, in Belgium, such a service has not yet come to birth, partly because of the difficulties raised by our complex constitutional organisation.

The project will aim at the drafting of a report presenting and evaluating the different scenarios of creation of a data archiving centre for the social sciences. The report should provide an answer to the following questions:

- What kind of data are to be gathered by a social archiving service ?
- What kind of metadata will be necessary to adequately describe these data ?
- What steps should be taken in order to create such an archiving service ?
- What are the legal and institutional restraints that could hamper the creation of such an archiving service ?
- Under what legal status should that service be created ?
- What is the role to be played by existing archiving institutions in the creation of the service?

The work is structured in different tasks to realise

Prospectstudie SODA  
Haalbaarheid van een data-archief voor de sociale wetenschappen

Task	Description
1	Typology of the bases and the data
2	Analyze of the needs of administrations, research centres and users of social sciences data
3	Analyze of similar projects
4	Interviews of potentially interested institutions
5	Analyze of institutional and legal framework
6	Analyse of techniques and procedures to archive documents
7	Analyse of the question of the use of data bases by thirds
8	Final report/ratio

Different steps have been or will be taken in order to reach that goal :

1. A study of foreign archiving centres considered as "best practices", via a desk analysis of their websites and a personal contact with their managers. We have selected following archiving centers : "UK Data Archive" in United Kingdom, "Zentralarchiv für Empirische Sozialforschung" in Germany, "Belgian Archive for the Social Sciences, and Council of European Social Science Data Archives.
2. A description of the needs of administrations, research centres and users of social sciences data is to be made soon through a limited number of group interviews;
3. Interviews with potentially interested institutions will take place in order to examine their possible contribution to such an initiative;
4. An international seminar will be organized to gather international experts in the field and compare their experiences.

### Research team

This research will therefore provide a specific proposal about the manner of conceiving such a service, taking into account the specific context of Belgium. For that purpose, different services specialized in the relevant fields of expertise have been teamed up:

1. **The Expertisecentrum e-David:** <http://www.expertisecentrumdavid.be/eng/index.php>, represented by Ms. Inge SCHOUPS.  
Expertisecentrum DAVID is a centre of research and knowledge on digital archiving. Its main contribution in the project is on the technical issues. The centre is working on the implication of the development of this portal. It is also interested in the techniques and procedures to archive documents.
2. **The Technology Assessment Research Unit from the University of Namur (CITA):** <http://www.fundp.ac.be/cita>, represented by Prof. Claire LOBET-MARIS and Ms. Véronique LAURENT.  
This research centre is specialised in the assessment of new technologies and more precisely, of information and communication technologies (ICT). Its main implication in the project will be on the content of the portal and on the analysis of the needs and requirements of researchers and upcoming users.
3. **The Centre for Computer and Law from the University of Namur (CRID):** <http://www.crid.be>, represented by Prof. Yves POULLET and Ms. Nathalie LEFEVER.

This centre is specialised in the study of legal problems and challenges linked to ICT. It will take in charge the legal questions linked to this portal (data protection, intellectual property rights, constitutional concerns, use of the data bases by third parties,...).

For more precisions about the project, members of the research team can be contacted:

1. **Inge Schoups**, Expertisecentrum DAVID, [inge@edavid.be](mailto:inge@edavid.be)
2. **Prof. Claire Lobet-Maris, Véronique Laurent**, Cellule Interdisciplinaire de Technology Assesement FUNDP, [clo@info.fundp.ac.be](mailto:clo@info.fundp.ac.be), [vla@info.fundp.ac.be](mailto:vla@info.fundp.ac.be)
3. **Prof. Yves Pouillet, Nathalie Lefever**, Centre de Recherche Informatique et Droit FUNDP, [yves.pouillet@fundp.ac.be](mailto:yves.pouillet@fundp.ac.be), [nathalie.lefever@fundp.ac.be](mailto:nathalie.lefever@fundp.ac.be)

## BIJLAGE B. Bibliografie

---

### LEGISLATION

Arrêté royal du 12 décembre 1957 concernant l'exécution de la loi du 24 juin 1955 relative aux archives, *M.B.* 20 décembre 1957

Arrêté royal du 13 février 2001 portant exécution de la loi du 8 décembre 1992 relative à la protection de la vie privée à l'égard des traitements de données à caractère personnel, *M.B.* 13 mars 2001

CESSDA, *Articles of governance*, disponible à l'adresse <http://extweb3.nsd.uib.no/opencms7final/export/sites/default/cessda/pdf/cessdaconstitution20040402.pdf>

CESSDA, *Trans-border Data Access Agreement*, disponible à l'adresse [http://extweb3.nsd.uib.no/opencms7final/export/sites/default/cessda/pdf/tbaa\\_v23.pdf](http://extweb3.nsd.uib.no/opencms7final/export/sites/default/cessda/pdf/tbaa_v23.pdf)

Commission Européenne, *Communication de la Commission au Parlement européen, au Conseil et au Comité économique et social européen sur l'information scientifique à l'ère numérique : accès, diffusion et préservation*, COM(2007) 56 final, 14 février 2007, disponible à l'adresse [http://ec.europa.eu/research/science-society/document\\_library/pdf\\_06/communication-022007\\_fr.pdf](http://ec.europa.eu/research/science-society/document_library/pdf_06/communication-022007_fr.pdf)

*Déclaration de Berlin sur le Libre Accès à la Connaissance en Sciences exactes, Sciences de la vie, Sciences humaines et sociales*, disponible à l'adresse [http://oa.mpg.de/openaccess-berlin/BerlinDeclaration\\_wsis\\_fr.pdf](http://oa.mpg.de/openaccess-berlin/BerlinDeclaration_wsis_fr.pdf)

Directive 95/46/CE du 24 octobre 1995 du Parlement européen et du Conseil relative à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, *J.O.C.E.* n° L281 du 23 novembre 1995, pp. 0031-0050.



Directive 96/9/CE du Parlement européen et du Conseil, du 11 mars 1996, concernant la protection juridique des bases de données, *J.O.* n°L077 du 27/03/1996, p.00-20 – 00-28

Loi du 11 avril 1994 sur la transparence administrative, *M.B.* 30 juin 1994

Loi du 24 juin 1955 relative aux archives, *M.B.* 12 août 1955

Loi du 30 juin 1994 relative au droit d'auteur et aux droits voisins, *M.B.* 27 juillet 1994

Loi du 8 décembre 1992 relative à la protection de la vie privée à l'égard des traitements de données à caractère personnel, *M.B.* 18 mars 1993

## DOCUMENTATIE

*ACROSS THE DECADES - 40 years of data archiving*, brochure distribuée par l'UKDA à l'occasion de ses 40 ans, dont le contenu est disponible en ligne à l'adresse <http://www.data-archive.ac.uk/ukda40/>

ASHLEY K., 'Digital archive costs: Facts and fallacies', in: *Proceedings of the DLM-Forum on electronic records. European citizens and electronic information: the memory of the Information Society. Brussels, 18-19 October 1999* (Luxemburg 2000) 121-126. Op WWW: <[http://europa.eu.int/ISPO/dlm/dlm99/dlm\\_proceed99\\_03.pdf](http://europa.eu.int/ISPO/dlm/dlm99/dlm_proceed99_03.pdf)>

ASHLEY K., 'Producing practical preservation procedures', in: *Proceedings of the DLMForum 2002. @ccess and preservation of electronical information: best practices and solutions* (Luxemburg 2002) 104-113

BEAGRIE N. en D. GREENSTEIN, *A strategic policy framework for creating and preserving digital collections*. British Library Research and Innovation Report 107 (Londen 1998)

BEARMAN D., 'Reality and chimeras in the preservation of electronic records', *D-Lib Magazine* (Volume 5 Number 4, April 1999). <http://www.dlib.org/dlib/april99/bearman/04bearman.html>

BLANK, G., & RASMUSSEN, K. B. (2004). The data documentation initiative: The value and significance of a worldwide standard. *Social Science Computer Review*, 22, 307--318. <http://www.icpsr.umich.edu/DDI/>

CHAPMAN S., 'Counting the costs of digital preservation: is repository storage affordable?'. *Journal of Digital Information* 4 (2003), issue 2, article 178. Op WWW: <<http://jodi.tamu.edu/Articles/v04/i02/Chapman/chapman-final.pdf>>

COX R.J., *Managing records as evidence and information* (Westport, Connecticut en Londen 2001)

DEPOORTERE, R., « Les archives en Belgique : une réalité éclatée », *Archives*, Vol. 34, Numéros 1 et 2, 2002-2003, p.101, disponible à l'adresse [http://www.archivistes.qc.ca/revuearchives/vol34\\_1-2/34-1&2-Depoortere.pdf](http://www.archivistes.qc.ca/revuearchives/vol34_1-2/34-1&2-Depoortere.pdf)

DE TERWANGNE, C., LOUVEAUX, S., « Protection de la vie privée face au traitement de données à caractère personnel : le nouvel arrêté royal », *J.T.*, 2001, p.457

DOLLAR Ch. M., *Archival theory and information technologies. The impact of information technologies on archival practices and methods* (Universiteit van Macerata, Italië 1992)

DOLLAR Ch. M., *Authentic electronic records: strategies for long-term access* (Chicago 1999)

DOORN P.K. en H.D. TJALSMA: 'Historical data archives: preserving and documenting historical data', in: *INSAR supplement II. The proceedings of the DLM-Forum on electronic records. Brussels 18-20 December 1996* (Luxemburg 1997) 155-160

DUMONT V., DE PAUW E., LAURENT V., LOBET-MARIS C., PONSAERS P., *Une bibliothèque virtuelle pour la Police Fédérale : analyse de la faisabilité du projet et cahier des charges méthodologique de l'application*, Rapport de recherche du projet "Agora-Bibliothèque Virtuelle", 2005 (Version électronique disponible sur [http://www.fundp.ac.be/recherche/publications/page\\_view/56910/](http://www.fundp.ac.be/recherche/publications/page_view/56910/))

*Economic and Social Data Service – Annual report 2000-2001*, disponible à l'adresse <http://www.data-archive.ac.uk/news/publications/AnnRep00.pdf>

*Economic and Social Data Service – Annual report 2003-2004*, disponible à l'adresse <http://www.data-archive.ac.uk/news/publications/annrep2004.pdf>

*Economic and Social Data Service - Strategic plan 2004-2009*, disponible à l'adresse <http://www.data-archive.ac.uk/news/publications/strategicplan0409.pdf>

HEDSTROM., 'Electronic archives: Integrity and access in the network environment', in: S. Kenna en S. Ross, *Networking in the Humanities* (Londen, etc. 1995) 77-95  
Jones M. en N. Beagrie, *Preservation management of digital materials. A handbook* (Londen 2001)

Inter-university Consortium for Political and Social Research (2005). Guide to social science data preparation and archiving (3rd ed.). Ann Arbor, Michigan.  
<http://www.icpsr.umich.edu/access/dataprep.pdf>

LAZINGER S.S., *Digital preservation and metadata: history, theory, practice* (Englewood, Colorado 2001)

LEONARD, T. et POULLET, Y., « La protection des données à caractère personnel en pleine (r)évolution », *J.T.*, 1999, p. 377, n°30

MOHLER, P.Ph., UHER, R. (2003): Documenting Comparative Surveys for Secondary Analysis. In: Harkness, Janet A.; Van de Vijver, Fons J.R.; Mohler, Peter Ph. (eds.): *Cross- Cultural Survey Methods*. Hoboken, New Jersey: Wiley.

Norme documentaire ISO 15836 :2003(E) : « Information and documentation – The Dublin Core metadata element set », 26 février 2003, document PDF téléchargé sur le site <http://www.niso.org/international/SC4/n515.pdf>

OECD Follow Up Group on Issues of Access to Publicly Funded Research Data (2003, March). Promoting access to public research data for scientific, economic, and social development. Final Report.

[http://dataaccess.ucsd.edu/Final\\_Report\\_2003.pdf](http://dataaccess.ucsd.edu/Final_Report_2003.pdf)

ROKKAN, S. (1964). Archives for secondary analysis of sample survey data: An early inquiry into the prospects of Western Europe. In: International Social Science Journal, 16, 49--62

ROKKAN, S. (1966). Data archives for the social sciences. Tools and methods of comparative research (Publications of the International Social Science Council, Vol. 3). Paris/La Hague: Mouton.

ROTHENBERG J. en BIKSON T., *Digital preservation: carrying authentic, understandable and usable digital records through time. report to the dutch national archives and ministry of the interior*, 1999 ([http://www.digitaleduurzaamheid.nl/bibliotheek/docs/final-report\\_4.pdf](http://www.digitaleduurzaamheid.nl/bibliotheek/docs/final-report_4.pdf))

ROTHENBERG J., *An experiment in using emulation to preserve digital publications*, Den Haag, 2000 (<http://www.kb.nl/coop/nedlib/results/emulationpreservationreport.pdf>)

ROTHENBERG J., *Avoiding technological quicksand: finding a viable technical foundation for digital preservation. a report to the council on library and information resources*, Washington, 1999 (<http://www.clir.org/pubs/reports/rothenberg/pub77.pdf>)

ROTHENBERG J., *Ensuring the longevity of digital information*, Santa Monica, 1999 (<http://www.clir.org/pubs/archives/ensuring.pdf>)

TJALSMA, H., *Archiveren van Digitaal Academisch Erfgoed: een verslag als voorbeeld*. DANS studies in digital archiving, Den Haag, 2006, 70 p.

SCHEUCH, E.K. (2003): History and visions in the development of data services for the social sciences. In: International Social Science Journal, 177: 385-399.

SCHURER K., *Better access to electronic information for the citizen. The relationship between public administration and archives services concerning electronic documents and records management* (Luxemburg 2001).

SHEPHERD E. en Ch. SMITH, 'The Application of ISAD(G) to the description of archival datasets', in: *Journal of the Society of Archivists*, 21 (2000), no 1, 55- 86

Te Royal Statistical Society & the UK Data Archive (2002). Preserving & sharing statistical material. Essex, Colchester.

<http://www.data-archive.ac.uk/news/publications/PreservingSharing.pdf>

## SITES INTERNET DE BASES DE DONNEES

European Social Survey  
<http://www.europeansocialsurvey.org>

International Federation of Data Organization for the social science  
<http://www.ifdo.org/>

Joint Information Systems Committee  
<http://www.jisc.ac.uk/>

Panel de Démographie Familiale, Université de Liège  
<http://www.psbh.ulg.ac.be/pages/donnees.htm>

## **ANNEXE 1. Liste de centres et de portails d'archives de données en sciences sociales**

---

Cette liste rassemble les principaux centres d'archive en sciences sociales en Europe et au niveau international mais est non exhaustive.

### **1. AU NIVEAU NATIONAL**

Belgique : le "Belgian Archives for the Social Sciences" - B.A.S.S :  
<http://bass.rspo.ucl.ac.be/>

Allemagne : le "Zentralarchiv für Empirische Sozialforschung", archives centrales de la recherche sociale empirique:  
<http://www.gesis.org/en/za/index.htm>

France: le Réseau Quetelet qui associe des centres français ayant des missions nationales en matière de diffusion et d'archivage d'ensemble de données statistiques pour les sciences humaines et sociales:  
<http://www.centre.quetelet.cnrs.fr/>

Suisse: le SIDOS, Service suisse d'information et d'archivage de données pour les sciences sociales:  
<http://www.sidos.ch/index.html>

Pays-Bas: le DANS, Data Archiving and Networked Services:  
<http://www.dans.knaw.nl/>

Finlande: le Finnish Social Science Data Archive:  
<http://www.fsd.uta.fi/english/>

Grèce: la Greek Social Databank:  
<http://www.gsdb.gr/>

Hongrie: le Tárki Data Archive:  
<http://www.tarki.hu/en/services/da/index.html>

Autriche : le WISDOM, Wiener Institut für Sozialwissenschaftliche Dokumentation und Methodik  
<http://www.wisdom.at/Default.aspx> - en allemand uniquement

République Tchèque: le Sociological Data Archive  
<http://archiv.soc.cas.cz/en/>

Danemark: le Danish Data Archive:  
<http://archiv.soc.cas.cz/en/>

Estonie: le Estonian Social Science Data Archive:  
<http://psych.ut.ee/esta/>

Irlande: le Irish Social Science Data Archive:  
<http://www.ucd.ie/issda/>

Norvège: les Norwegian Social Science Data Services:  
<http://www.nsd.uib.no/english/>

Roumanie: le Romanian Social Data Archive:  
<http://www.roda.ro/>

Slovénie: ADP  
<http://www.adp.fdv.uni-lj.si/en/prva.htm>

Suède: Swedish Social Science Data Service:  
<http://www.ssd.gu.se/enghome.html>

Grande-Bretagne: UK Data Archive:  
<http://www.data-archive.ac.uk/>

## 2. AU NIVEAU INTERNATIONAL :

Europe: le Council of European Social Science Data Archives: <http://www.nsd.uib.no/cessda/>

Europe de l'Est: l'EDAN, Est European Data Archive Network:  
[http://www.gesis.org/en/cooperation/data\\_service/eastern\\_europe/index.htm](http://www.gesis.org/en/cooperation/data_service/eastern_europe/index.htm)

Réseau internuniversitaire: le Inter-university Consortium for Political and Social Research,  
regroupant +/- 500 universités:  
<http://www.icpsr.umich.edu/>

International Federation of Data Organization for the social science  
<http://www.ifdo.org/>

## ANNEXE 2. Grille d'analyse sociologique des portails d'archives en sciences sociales

Portails cibles

Europe: le Council of European Social Science Data Archives:  
<http://www.nsd.uib.no/cessda/>

Grande-Bretagne: UK Data Archive:  
<http://www.data-archive.ac.uk/>

Allemagne : le "Zentralarchiv für Empirische Sozialforschung", archives centrales de la recherche sociale empirique:  
<http://www.gesis.org/en/za/index.htm>

Pays-Bas: le DANS, Data Archiving and Networked Services:  
<http://www.dans.knaw.nl/>

Belgique : le "Belgian Archives for the Social Sciences" - B.A.S.S. -  
<http://logi.rspo.ucl.ac.be/>

Critères d'analyse	Détails
<b>Type d'application</b>	<ul style="list-style-type: none"> <li>➤ Bibliothèque virtuelle (accès possible en ligne aux documents via le portail)</li> <li>➤ Portail d'information (portail recensant et redirigeant des ressources existant par ailleurs)</li> <li>➤ Référothèque (référence de sources uniquement, sans forcément de liens directs vers ces sources)</li> </ul>
<b>But de l'application</b>	A quels besoins ce portail vise-t'il à répondre ?
<b>Public cible de l'application</b>	Utilisateurs potentiels
<b>Données dans les portails</b>	<ul style="list-style-type: none"> <li>➤ Type de données : quantitatives – qualitatives</li> <li>➤ Disciplines</li> <li>➤ Thèmes généraux autour du portail (consommation, vote, ...)</li> <li>➤ Format des données (audio, vidéo)</li> <li>➤ Système d'exploitation des données (Oracle, SAS, ...)</li> <li>➤ Nature des données : rapports de recherche (quels types de rapports de recherche et quelles type de recherche ?). Les données exposées sont-elles étendues aux rapports de recherche ? (livres, articles, revues, ...)</li> </ul>
<b>Fonctionnalités du portail (options)</b>	<ul style="list-style-type: none"> <li>➤ Présentation du moteur de recherche (recherche simple, avancée, personnalisée, thésaurus ?)</li> <li>➤ Personnalisation du profil chez l'utilisateur</li> <li>➤ Sauvegarder ses recherches</li> <li>➤ Newsletter</li> <li>➤ Liens externes vers d'autres sites</li> </ul>

Prospectstudie SODA  
Haalbaarheid van een data-archief voor de sociale wetenschappen

<b>Présentation des sources documentaires</b>	<ul style="list-style-type: none"> <li>➤ Mode de présentation des données (fiches documentaires, listes de ressources, ...)</li> <li>➤ Types de méta-données utilisées</li> </ul>
<b>Accès au portail</b>	<ul style="list-style-type: none"> <li>➤ Accès aux utilisateurs/aux données restreint ou non restreint</li> <li>➤ Accès restreint pour certaines personnes (qui)</li> <li>➤ Moyens d'accès au portail : authentification nécessaire (login-mot de passe) ou libre consultation</li> </ul>
<b>Gestion du portail</b>	<ul style="list-style-type: none"> <li>➤ Quels sont les gestionnaires du portail ?</li> <li>➤ Cette gestion est-elle centralisée ou décentralisée ?</li> <li>➤ Le portail est-il statique ou dynamique (possibilité que le portail soit alimentés par des utilisateurs)</li> <li>➤ Y a-t-il des rôles distribués au sein de l'application ? (rôles à titre d'exemples : lecteur, rédacteur, validateur, administrateur, webmaster)</li> <li>➤ Règles de gestion (sur le contenu, sur les statuts) dans l'application</li> </ul>



### **ANNEXE 3. Liste de questions juridiques pour l'établissement d'une grille d'analyse des portails en sciences sociales**

---

1. Quel est le statut juridique de la plateforme ?
2. Quel est le rôle de la plateforme ?
  - publication des données ?
  - établissement d'un répertoire des données disponibles ?
  - vérification de la légalité de la collecte originale des données ?
  - codage des données ?
  - autres ?
3. Quelles sont les obligations mises à charge de la plateforme ?
  - sécurité des données ?
  - information des personnes concernées ?
  - octroi des droits d'accès ?
  - déclaration de traitement des données ?
  - autres ?

Si la plateforme ne prend pas ces obligations en charge, qui le fait ?
4. Quelles sont les conditions de collecte de données par la plateforme ?
  - Quelles sont les caractéristiques des données collectées : personnelles/codées/anonymes ?
  - Quelles sont les sources de données ?
  - Quels sont les critères qui vont décider de la collecte ou non ?
  - Y a-t-il un contrat entre l'auteur original de la base de données et la plateforme ?
  - Quels sont les droits/contreparties accordés à l'auteur de la base de données ?
  - Quels sont les droits/contreparties accordés à l'employeur de l'auteur ?
  - Quels sont les droits/contreparties accordés à celui qui finance la base de données ?
  - Sous quelles conditions les données sont-elles mises à disposition de la plateforme ?
  - Sous quelles conditions la plateforme échange-t-elle des données avec l'étranger ?
5. Quelles sont les conditions d'accès à la plateforme ?
  - Y a-t-il un contrat entre la plateforme et le ré-utilisateur ?
  - Quelles sont les conditions concernant le ré-utilisateur ?
  - Des obligations sont-elles mises à charge du ré-utilisateur ?
  - Quelles sont les types d'accès autorisés : réutilisation/copie, accès total/partiel,... ?
  - Y a-t-il un mécanisme garantissant un accès limité aux données nécessaires, et si oui, lequel ?
6. Pour quels types de réutilisation les données sont-elles disponibles : recherche scientifique ? utilisation commerciale ?
7. L'œuvre issue de la réutilisation est-elle protégée contre les prétentions en matière de droits d'auteur de celui qui a récolté les données ?
8. (pour l'Allemagne) Comment se règlent les questions constitutionnelles liées à l'établissement de la plateforme :
  - S'agit-il d'une plateforme créée au niveau fédéral/fédéré/privé ?
  - Sous quels termes les recherches publiques sont-elles mises à la disposition de la plateforme ?

## **ANNEXE 4. Exemple de questionnaire utilisé pour les visites des centres d'archives**

---

### **Topics of interest about the Zentralarchiv für empirische Sozialforschung (ZA)**

Thank you very much for taking time in reading the following questions.

They list our topics of interest.

We would like to discuss these points during our meeting of June 19

If you have time, you can also answer to part or the whole of this questionnaire in advance.  
Your answers will provide us with very important information, allowing us to analyze the possibility of successfully setting up a data archiving centre in Belgium on the ZA model.

To make the questionnaire as clear as possible, it has been divided into arbitrary sections. We have already analyzed carefully the information available on your website, you therefore do not need to repeat it, nor do you need to take into account the limit of space available for every answer.

Feel free to attach any document that might be useful for our understanding, such as official papers, diagrams, etc., or to add any information that might serve our purposes, even if it does not fit into the answers to our questions.

Would you need any precision on the scope of the questions, do not hesitate to contact us:

Véronique Laurent  
+32 81 72 52 49  
vla@info.fundp.ac.be  
<http://www.info.fundp.ac.be/cita>

### **Contact information**

***If you answer part or the whole of this questionnaire, thank you for leaving us your contact details, in case you would agree to give further explanation possibly needed:***

Organisation :

Country :

First Name :

Second Name :

Department :

E-mail :

Phone number :

### **Creation of ZA**

What are the major reasons and events that have motivated the setting of ZA?

Was the creation of ZA an initiative of the States or of the Federal Government ? Does it have a federal statute? Which one?

What were the initial ambitions of the ZA? Have they evolved along the time?

Is it possible to get a copy of the charters ruling the ZA ?

What were the main difficulties in the creation and the development of ZA? *(give 3 main)*

What were the main facilitators for the creation and the development of ZA? *(give 3 main)*

### **Missions and targets of ZA**

What are the main missions and activities of ZA? Have they evolved along the time?

- *In your website, it is quoted "The ZA provides resource discovery and support for secondary use of quantitative and qualitative data in research, teaching and learning" (...)*
- *How long, and why has ZA been involved in research and development activities ?*

What definition does the centre assert as "empirical social research"? (by reference to the collection of digital data of ZA) What are the main differences with the collection of the social science information center (IZ)?

- *Thanks to explain as much as possible your definition, what authorities, references, schools, are used...*

Why is your centre particularly interested in "empirical social research" for its collection?

### **Organisation, finance and resources**

#### **Organisation**

How is the UK Data Archive organisation structured? Is it possible to get an organisation chart of ZA ?

What are the roles of the departments?

What are the functions of the employees? *(please fill in this table with percentages)*

<b>Functions</b>	<b>Percentage in ZA</b>
Scientists/Academics	... %
Administrative officers	... %
Statisticians	... %
Computer scientists	... %
Other functions (please give more details)	... %
- .....	...%
- .....	...%
- .....	...%
- .....	...%
- .....	...%
- .....	...%
- .....	...%
- .....	...%

How many people are currently working for ZA? (in "equivalent full time") .... EFT

Who manages the portal : <http://www.gesis.org/en/iz/index.htm> and how does it work?

Does ZA take in charge the validation of contents and their compliance with the Law (notably the DP Act)? Have you already incurred any suit in that context?

Who is responsible for this validation?

## Finance and resources

What is the annual average budget of ZA ? Has that budget evolved (increased or decreased) over the last ten years and why?

Could you give the details of the budget? What is the part of each cost (staff, hardware infrastructure, maintenance, other)?

What are the main funding sources that support ZA? *(please fill in this table with percentages)*

Sources of funding	Percentages of funding in ZA
Universities	... %
German Federal State	... %
German States	... %
Cologne Association for Social Research	... %
German Social Science Infrastructure Service	... %
Self funding	... %
Other sources of funding (please give more details)	... %
- .....	...%
- .....	...%
- .....	...%
- .....	...%
- .....	...%
- .....	...%
- .....	...%

## Partnership or membership

What are the main partners of ZA ? Could you briefly describe their roles in the partnership? *(please fill in this table)*. Were there modifications in the partnerships during the last ten years worth being underlined ?

Main partners	Roles

- Does ZA regularly work with Social Science Information Center (IZ) and with the Center for Survey Research and Methodology (ZUMA)?
- How does the cooperation with the University of Cologne actually work ?

What are the main organisations of which ZA is a member? Could you briefly describe their roles in the partnership? *(please fill in this table)*

Main organisations	Roles

Prospectstudie SODA  
Haalbaarheid van een data-archief voor de sociale wetenschappen


Did ZA conclude formal partnership agreements with official agencies, international agencies, universities, departments and/or individual academics,... concerning the availability of their raw-data or computer-readable datasets ?

- What are the conditions of these agreements?
- Could you send us an example of these agreements?
- If not, how are you working with your partners?

### **Data supply and use in ZA**

Have you got some statistical figures about the users, depositors and data creators at the ZA?  
(percentages, numbers, ...)

- Is it possible to obtain a copy of them ?

### **Suppliers**

What are the main data suppliers, offering their data to ZA ?  
(Please fill in this table with percentages and precise if the datasets provided by these data-suppliers have decreased or increased over the last ten years)

<b>Data-suppliers</b>	<b>Percentages of datasets in ZA</b>	<b>Evolution:</b> increasing ("up"), decreasing ("down"), or constant ("constant")
Official Agencies	... %	
International Statistical Agencies	... %	
Individual Academics	... %	
Market research agencies	... %	
Historical agencies	... %	
Other data suppliers (please give more details)	... %	
- .....	...%	
- .....	...%	
- .....	...%	
- .....	...%	
- .....	...%	
- .....	...%	
- .....	...%	

### **Data supplied**

What types of data, services and subjects are mainly supplied? (please fill in this table with percentages and indicate the evolution of those categories of incomes over the last ten years)

<b>Data supplied</b>	<b>Percentages</b>	<b>Evolution:</b> increasing (up), decreasing (down), or constant

Prospectstudie SODA  
Haalbaarheid van een data-archief voor de sociale wetenschappen

		<b>(constant)</b>
<u>Types of data</u> Quantitative Qualitative Multimedia Non-digital material Other types (please give more details) - ..... - ..... - ..... - ..... - ..... - ..... - .....	<b>(Tot: 100%)</b> ... % ... % ... % ... % ... % ... % ... % ... % ... % ... %	("up", "down", "constant")
<u>Study collections</u> <b>German Study Collections</b> ALLBUS GDR – new federal states Politbarometer Elections studies IAB – Sample Survey of the Employed  <b>International Study Collections</b> Eurobarometer ISSP (International Social Survey Programme) Eastern Europe	<b>(Tot: 100%)</b> ...% ... % ... % ... % ... % ... % ... %  ...% ... % ... % ... %	("up", "down", "constant")
<u>Main supplied topics</u> <i>(state, political ideology, social policy, professions, family, income ...)</i> (please give more details) - ..... - ..... - ..... - ..... - ..... - ..... - .....	<b>(Tot: 100%)</b> ...% ...% ...% ...% ...% ...% ...%	("up", "down", "constant")

## Releasing of the data

Is there a selection made by the ZA in the data to be collected?

- *On what basis is that selection made?*
- *For what reasons are you making this selection (Quality of the data? Privacy questions ? or...?)*

Is there a formal agreement between the data suppliers and the Z.A? Are there specific provisions regarding the quality of the data, their format, their potential use? Is it possible to obtain a copy of this agreement ?

What kind of data do you store: anonymous, coded, personal data ?

- *In case of anonymized data, who is in charge of anonymization ?*
- *Does ZA take in charge the obligations related to privacy (verification of the rightful collection of data, consent for reuse, information,...) in case of personal data ?*

What are the main requirement addressed to the suppliers regarding the format, the metadata and the quality of their data ?

- *Format?*

- Metadata ?
- Quality ?

## Users

What are the main users, accessing to data in ZA ?

(please fill in this table with percentages and precise if the number of these users is increasing or decreasing over the last ten years )

Users	Percentages of users of ZA	Evolution : increasing (up), decreasing (down), or constant (constant)
	(Tot: 100%)	("up", "down", "constant")
Scientifics and academics	... %	
Political public	... %	
Public administration	... %	
Business public	... %	
Medias	... %	
Civil society	... %	
Other users (please give more details)	... %	
- .....	...%	
- .....	...%	
- .....	...%	
- .....	...%	
- .....	...%	
- .....	...%	
- .....	...%	

What is the part of domestic users and foreign users accessing to data in ZA? Has that repartition evolved over the ten last years?

Users	Percentage	Evolution : increasing (up), decreasing (down), or constant (constant)
Domestic users	...%	
Foreign users	...%	

What types of data, services and subjects are mainly browsed or consulted? (please fill in this table with percentages and precise if these data are increasing or decreasing over the last ten years)

Use of Data	Percentages	Evolution : increasing (up), decreasing (down), or constant (constant)
Types of data	(Tot: 100%)	("up", "down", "constant")
Quantitative	... %	
Qualitative	... %	
Multimedia	... %	
Non-digital material	... %	
Other type	... %	

<p>(please give more details)</p> <p>- ..... %</p> <p>- ..... %</p> <p>- ..... %</p> <p>- ..... %</p> <p>- ..... %</p> <p>- ..... %</p> <p>- ..... %</p>	<p>... %</p> <p>... %</p> <p>... %</p> <p>... %</p> <p>... %</p> <p>... %</p> <p>... %</p>	
<p><u>Study collections</u></p> <p><b>German Study Collections</b></p> <p>ALLBUS</p> <p>GDR – new federal states</p> <p>Politbarometer</p> <p>Elections studies</p> <p>IAB – Sample Survey of the Employed</p> <p><b>International Study Collections</b></p> <p>Eurobarometer</p> <p>ISSP (International Social Survey Programme)</p> <p>Eastern Europe</p>	<p><b>(Tot: 100%)</b></p> <p>...%</p> <p>... %</p> <p>... %</p> <p>... %</p> <p>... %</p> <p>... %</p> <p>...%</p> <p>... %</p> <p>... %</p> <p>... %</p>	<p>("up", "down", "constant")</p>
<p><u>Main consulted catalogues</u></p> <p>ZA Data Holding Catalogue</p> <p>ALLBUS</p> <p>Historical Statistics Online</p> <p>ZA library 'Empirische Sozialforschung'</p> <p>International Federation of Data Organizations</p> <p>Others types of catalogues</p> <p>(please give more details)</p> <p>- ..... %</p> <p>- ..... %</p> <p>- ..... %</p> <p>- ..... %</p>	<p><b>(Tot: 100%)</b></p> <p>... %</p> <p>... %</p> <p>... %</p> <p>... %</p> <p>... %</p> <p>... %</p> <p>... %</p> <p>... %</p> <p>... %</p> <p>... %</p>	<p>("up", "down", "constant")</p>
<p><u>Main supplied topics</u></p> <p><i>(state, political ideology, social policy, professions, family, income ...)</i></p> <p>(please give more details)</p> <p>- ..... %</p> <p>- ..... %</p> <p>- ..... %</p> <p>- ..... %</p> <p>- ..... %</p> <p>- ..... %</p> <p>- ..... %</p>	<p><b>(Tot: 100%)</b></p> <p>...%</p> <p>...%</p> <p>...%</p> <p>...%</p> <p>...%</p> <p>...%</p> <p>...%</p>	<p>("up", "down", "constant")</p>

What types of data are in "open-access" (opportunity to access to the documents without any registration) ?

## ZA regulations of use

The website of ZA features “regulations of use” applying to every user; does the procedure of request for material imply a formal acceptance of these regulations?

These “regulations of use” mention different categories of users;  
- who decides which category a user belongs to ?



- who decides which category a set of data belongs to ?

The "regulations of use" mention an "international archive convention"

- *"In the context of the international archive convention, users living outside Germany are asked to direct their requests regarding data to their home archive first"*

Could you give us more details about this convention?

Do the "regulations of use" contain all the requirements addressed to the data users regarding the sources citation, the data use,... ? If not, what other main requirements ought to be mentioned ?

Is the ZA bound by any particular legislation when disseminating the data you have collected (e.g. Act on the exploitation of data collected by the public bodies or privacy legislation?)

Are you taking certain security measures for transmitting the data ?

Is ZA involved in deciding on the copyrights on the scientific works based on reused data?

### **General opinion about ZA**

What are the main advantages you can lay down from ZA?

What are the main inconveniencies you can lay down from ZA ?

What are the main lessons you can lay down from your experience?

What are the main recommendations you can give for the data archive project in Belgium ?

Have you got perspectives for the future of this portal?

Are there certain connections with similar data bases established at the local or federate states' levels?

### **Conclusion**

Have you other information to give to us?

**Thank you so much for having answered these questions!**

## **ANNEXE 5. Programme de discussion du groupe institutionnel (27 juin 2007)**

---

9h30 – 9h45	Accueil des participants et présentation du projet	
9h45 – 10h15	Un service d'archivage, dans quels buts ? Motivations et apports de création d'un service d'archivage de données en sciences sociales	Claire Lobet
10h15 -10h30	Pause café	
10h30 – 11h20	Comment ce service peut-il se présenter ? Présentation et fonctionnalités de ce service d'archivage	Véronique Laurent Nathalie Lefever
11h20 – 12h	Comment mettre ce service en place ? Implication institutionnelle et gestion autour de ce service d'archivage	Yves Poullet
12h	Lunch	

Nom : .....  
Institution : .....

Phase 1 : Motivations et apports de la création d'un service d'archivage de données en sciences sociales 9h45 – 10h15
--

### Les motivations pour la création d'un service d'archivage en sciences sociales en Belgique

- Quelles sont les principales motivations scientifiques à la création d'un tel service ?
- Quelles sont les principales motivations politiques à la création d'un tel service ?
- Quelles sont les principales motivations administratives à la création d'un tel service ?

.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....

### Les apports d'un service d'archivage en sciences sociales

- Quels sont les principaux apports de la création d'un tel service pour les chercheurs ?
- Quels sont les principaux apports de la création d'un tel service pour les administrations ?
- Quels sont les principaux apports de la création d'un tel service pour les politiques ?

.....  
.....  
.....  
.....  
.....  
.....  
.....

Phase 2 : Présentation et fonctionnalités de ce service d'archivage des données en sciences sociales  
10h30 – 11h20

Comment concevez-vous le futur service d'archivage Belge ?  
Nous vous proposons différents scénarios, non limitatifs.  
Vous pouvez cocher ceux que vous souhaitez.  
Plusieurs réponses peuvent être possibles...

### I. Missions principales du service d'archivage

- ☐ Conserver les données
- ☐ Favoriser leur accès et leur réutilisation
- ☐ Faciliter un échange de données entre institutions et chercheurs
- ☐ .....
- ☐ .....

### II. Types de données présentes dans le portail

#### Qualité des données

- ☐ Données quantitatives
- ☐ Données qualitatives
- ☐ Ressources multimédia
- ☐ Matériel non digital
- ☐ Autres types de données : .....

#### Disciplines

- ☐ Toutes les données se rapportant aux sciences sociales au sens large : sociologie, économie, psychologie, histoire, criminologie, démographie, statistique, autres, ...
- ☐ Uniquement la sociologie
- ☐ Uniquement des bases de données quantitatives, peu importe leur discipline
- ☐ .....

#### Caractère des données

- ☐ Données "brutes" (chiffres collectés, extraits d'interviews, ...)
- ☐ Données "travaillées" : rapports de recherche, articles, guides méthodologiques, questionnaires, déclaration de variables, plans de traitements, ... (documentation)
- ☐ Données personnelles
- ☐ Données anonymisées
- ☐ .....
- ☐ .....

### III. Fournisseurs de données dans le portail

- ☐ Agences de statistique officielles nationales : INS, iweps, ...
- ☐ Agences de statistique officielles internationales : eurobarometer...
- ☐ Universités, FNRS, enseignement supérieur
- ☐ Sociétés privées : SONECOM, ...
- ☐ Associations
- ☐ Presse et médias
- ☐ Etudiants
- ☐ Tout autre personne
- ☐ .....

#### IV. Utilisateurs potentiels de données

- ☐ Les chercheurs
- ☐ Les administrations de la recherche
- ☐ Le monde politique et les administrations publiques
- ☐ Entreprises
- ☐ Presse et médias
- ☐ Etudiants
- ☐ Tout autre personne
- ☐ .....

#### V. Accès aux données

##### Situation

- ☐ Portail accessible sur Internet
- ☐ Portail accessible sur Intranet
- ☐ Portail accessible des deux manières

##### Accès au contenu

- ☐ Accès directement aux données (virtual library)
- ☐ Système de liens vers des données existant sur d'autres sites (portail d'information)
- ☐ Système de référencement de données (référothèque)
- ☐ .....

##### Moyen d'accès aux données

- ☐ Accès libre aux données
- ☐ Accès aux données avec identification, login et mot de passe
- ☐ Accès aux données sur base d'un formulaire à remplir
- ☐ .....

##### Dans le cas d'un accès direct aux données

- ☐ Accès nécessitant de s'enregistrer pour accéder à l'ensemble du portail d'archivage
- ☐ Accès ouvert uniquement aux fiches documentaires des ressources et nécessité de s'enregistrer dans le système pour avoir accès aux données
- ☐ Accès ouvert pour les fiches documentaires et les documents travaillés, mais fermé pour les données brutes
- ☐ Accès totalement ouvert pour l'ensemble du portail d'archivage
- ☐ Fixation de l'accès aux données par le dépositaire du document dans le portail selon des catégories prédéfinies : ex : données accessibles à tout le monde, données uniquement réservées à l'enseignement et à la recherche, ...
- ☐ .....

##### Accès payant ou gratuit

- ☐ Accès gratuit aux données
- ☐ Accès payant aux données
- ☐ Formule combinée
- ☐ .....

#### VI. Fonctionnalités et services du portail

- ☐ Moteur de recherche : recherche simple ou recherche avancée (matière, thème, auteur, titre, date de la recherche, date de publication, type de donnée, ...)
- ☐ Liens vers des catalogues de bibliothèque
- ☐ Liens vers des conférences
- ☐ Liens vers des publications
- ☐ Nouveautés

- ☐ Outils de recherche ou d'analyse documentaires (convertisseur de données, moteurs de recherche, dictionnaires, ...)
- ☐ Mailing list
- ☐ "My space" : possibilité de sauvegarder ses recherches dans un espace personnel
- ☐ .....

## VII. Missions secondaires du service d'archivage

- ☐ Activités de Recherche et Développement dans le domaine de l'archivage (intégration de la recherche à l'archivage)
  - ☐ Conférences
  - ☐ Séminaires
  - ☐ Publications
  - ☐ .....
- ☐ Formation aux techniques d'archivage
- ☐ Envoi de supports documentaires
- ☐ Co-opérations avec d'autres centres d'archive et des bases de données internationales
- ☐ .....

Phase 3 : Mise en place concrète et implication par rapport à ce service d'archivage 11h20 – 12h00
---

Comment mettre en place ce portail ?
--------------------------------------

<u>Vous pouvez cocher ceux que vous souhaitez</u>
---

<u>Plusieurs scénarios et pistes de réflexion peuvent être possibles...</u>
---

## I. Gestion du service d'archivage

- ☐ Service créé et maintenu uniquement par l'Etat Fédéral (à la politique scientifique fédérale par exemple)
- ☐ Service créé et maintenu par les entités fédérale et fédérées
- ☐ Service créé et maintenu par l'Etat Fédéral en collaboration avec des universités (BASS)
- ☐ Service créé et maintenu par les entités fédérales, fédérées et les universités
- ☐ .....

## II. Sources de financement du service d'archivage

- ☐ Communauté Européenne
- ☐ Etat Fédéral
- ☐ Régions
- ☐ Communautés
- ☐ Auto-financement : cotisations institutionnelles, ....
- ☐ .....
- ☐ .....

## III. Implications des institutions, des centres de recherche et des fournisseurs de données dans le portail

- ☐ Un système de partenariat ?
- ☐ Sous quelle forme ?
- ☐ A quelles conditions ?
- ☐ Partenariat avec d'autres portails d'archives étrangers et internationaux
- ☐ Obstacles à la mise à disposition des données (juridiques, techniques, scientifiques, ...)

## IV. Comment alimenter le système et y convaincre des centres de recherche de déposer leurs données ? Il s'agit de trouver un processus "gagnant-gagnant" Comment attirer les chercheurs ?

- ☐ Chaque organisme ou chercheur qui fournit des données dans le portail ou s'y implique peut accéder, en contrepartie, aux données du service d'archive.
- ☐ Chaque organisme ou chercheur qui fournit des données dans le portail ou s'y implique peut recevoir gratuitement, en contrepartie, les données du portail sur des formats cd par exemple
- ☐ Tous les centres de recherches, ou les chercheurs qui ont des subventions des organismes subventionnant le service d'archivage (l'Etat fédéral, les communautés, les régions, ...) doivent, en contrepartie, mettre leurs données en ligne sur ce portail
- ☐ Les chercheurs et organismes qui publient dans le portail d'archive, sont en compensation, assurés de recevoir une "publicité" à leur recherche et leurs données...
- ☐ .....

#### V. Dynamique du portail d'archive ?

- ☐ Différents rôles et statuts possibles autour de l'application : lecteur, rédacteur, validateur, administrateur, webmaster, ...
- ☐ Gestion exclusive de l'application informatisée par le ou les organismes gérant le service d'archivage
- ☐ Gestion du portail d'archive par un ensemble d'institutions partenaires travaillant ou s'intéressant au domaine de l'archivage (comité de gestion)
- ☐ .....

#### VI. Divers

Comment assurer la pérennité des données mais aussi leur actualisation ?

Budget et coûts de ce service d'archive

Localisation du service d'archivage

- ☐ Au sein de la politique scientifique fédérale
- ☐ Dans l'entité en charge de sa gestion
- ☐ Dans une université
- ☐ Localisation mixte, partagée à plusieurs endroits
- ☐ .....

## **ANNEXE 6. Programme de discussion des représentants des universités et de la recherche (20 juillet 2007)**

9h30 – 9h45	Accueil des participants, tour de table et présentation du projet	Inge Schoups
9h45 – 10h00	Un service d'archivage, dans quels buts ? Motivations et apports de la création d'un service d'archivage de données en sciences sociales	Claire Lobet
10h00 – 10h45	Comment ce service peut-il se présenter ? Présentation et fonctionnalités de ce service d'archivage	Véronique Laurent
10h45 – 11h	Pause café	
11h - 11h40	Quelles seraient les conditions du dépôt des données ? Exigences des producteurs initiaux des données quant à leur dépôt dans le service d'archivage	Nathalie Lefever, Véronique Laurent
11h40 – 12h20	Quelles seraient les conditions de l'accès et de la réutilisation des données ? Vision des utilisateurs du service d'archivage quant à l'accès et la réutilisation des données	Véronique Laurent, Nathalie Lefever
12h20 – 12h30	Conclusion des débats et fin de la discussion	

Nom : .....  
Institution : .....  
Tel : .....  
E-mail : .....

### **Phase 1 : Motivations et apports de la création d'un service d'archivage de données en sciences sociales**

#### Les motivations pour la création d'un service d'archivage en sciences sociales en Belgique

- Quelles sont les principales motivations scientifiques à la création d'un tel service ?
- Quelles sont les principales motivations administratives à la création d'un tel service ?

.....  
.....  
.....  
.....  
.....  
.....  
.....

#### Les apports d'un service d'archivage en sciences sociales

- Quels sont les principaux apports de la création d'un tel service pour les chercheurs ?
- Quels sont les principaux apports de la création d'un tel service pour les universités / centres de recherche ?

Phase 2 : Présentation et fonctionnalités de ce service d'archivage des données en sciences sociales

Comment concevez-vous le futur service d'archivage belge ?  
*Nous vous proposons différents scénarios, non limitatifs.*  
*Vous pouvez cocher ceux que vous souhaitez.*  
*Plusieurs réponses peuvent être possibles et également complétées par d'autres réponses...*

#### I. Missions principales du service d'archivage

- ☐ Garantir une conservation sécurisée des données
- ☐ Favoriser leur accès et leur réutilisation
- ☐ Faciliter un échange de données entre institutions et chercheurs
- ☐ Réalisation d'une certification garantissant/reconnaissant la qualité des données déposées
- ☐ .....
- ☐ .....

#### II. Missions secondaires du service d'archivage

- ☐ Activités de Recherche et Développement dans le domaine de l'archivage (intégration de la recherche à l'archivage)
  - ☐ Conférences
  - ☐ Séminaires
  - ☐ Publications
  - ☐ .....
- ☐ Formation aux techniques d'archivage
- ☐ Envoi de supports documentaires
- ☐ Coopération avec d'autres centres d'archive et des bases de données internationales
- ☐ .....

#### III. Fonctionnalités et services du portail

- ☐ Moteur de recherche : recherche simple ou recherche avancée (matière, thème, auteur, titre, date de la recherche, date de publication, type de donnée, ...)
- ☐ Liens vers des catalogues de bibliothèque
- ☐ Liens vers des conférences
- ☐ Liens vers des publications
- ☐ Nouveautés
- ☐ Outils de recherche ou d'analyse documentaires (convertisseur de données, moteurs de recherche, dictionnaires, ...)
- ☐ Mailing list
- ☐ "My space" : possibilité de sauvegarder ses recherches dans un espace personnel
- ☐ .....
- ☐ .....

Phase 3 : Conditions du dépôt des données



11h – 11h40

Comment, en tant que producteur initial de données, concevez-vous les conditions de leur dépôt auprès du système d'archivage ?

*Nous vous proposons différents scénarios, non limitatifs.*

*Vous pouvez cocher ceux que vous souhaitez.*

*Plusieurs réponses peuvent être possibles et également complétées par d'autres réponses...*

I. Comment, à l'heure actuelle, conservez-vous les données issues des recherches terminées ?

- ☐ Pas de conservation
- ☐ Conservation assurée par le chercheur
- ☐ Conservation centralisée au niveau du centre de recherche / de l'université
- ☐ Conservation confiée à un centre d'archivage existant : .....  
.....  
.....
- ☐ Autre : .....

II. Types de données susceptibles d'être mises à la disposition du service d'archivage

Qualité des données

- ☐ Données quantitatives
- ☐ Données qualitatives
- ☐ Ressources multimédia
- ☐ Matériel non digital
- ☐ Autres types de données : .....

Disciplines

- ☐ Toutes les données se rapportant aux sciences sociales au sens large : sociologie, économie, psychologie, histoire, criminologie, démographie, statistique, autres, ...
- ☐ Uniquement la sociologie
- ☐ Uniquement des bases de données quantitatives, peu importe leur discipline
- ☐ .....
- ☐ .....

Caractère des données

- ☐ Données "brutes" (chiffres collectés, extraits d'interviews, ...)
- ☐ Données "travaillées" : rapports de recherche, articles, guides méthodologiques, questionnaires, déclaration de variables, plans de traitements, ... (documentation)
- ☐ Données personnelles
- ☐ Données anonymisées
- ☐ .....
- ☐ .....
- ☐ .....

III. Conditions favorisant la mise à disposition de données via le portail

- ☐ Accès limité en fonction du choix du déposant :
  - ☐ Choix entre plusieurs niveaux de disponibilité (basés sur des catégories d'utilisateurs / des types d'utilisation)
  - ☐ Choix au cas par cas : autorisation individuelle
  - ☐ Choix d'une limitation temporelle : pas disponible avant... ou disponible jusqu'à...

- ☐ Accès aux données du portail limité aux déposants
- ☐ Obligation pour le réutilisateur de rendre accessible les données dérivées
- ☐ Obligation de citation par le chercheur qui réutilise les données
- ☐ Droit d'utiliser ou de publier les données ailleurs
- ☐ Droit de retirer les données du service d'archive
- ☐ Convention générale entre le centre d'archive et les centres de recherche, universités, sources de financement,...
- ☐ Support dans la préparation des données pour l'archivage
- ☐ Publication et diffusion d'un répertoire de la recherche
- ☐ .....
- ☐ .....

#### IV. Fournisseurs dont vous souhaiteriez retrouver les données dans le portail

- ☐ Agences de statistique officielles nationales : INS, iweps, ...
- ☐ Agences de statistique officielles internationales : eurobarometer...
- ☐ Universités, FNRS, enseignement supérieur
- ☐ Sociétés privées : SONECOM, ...
- ☐ Associations
- ☐ Presse et médias
- ☐ Etudiants
- ☐ Toute autre personne
- ☐ .....
- ☐ .....

#### V. Possédez-vous des bases de données susceptibles d'être mises à la disposition du service d'archivage ?

.....  
.....  
.....  
.....  
.....  
.....

#### VI. Obstacles à la mise à disposition des données par le chercheur

- ☐ Contrat de financement transférant la propriété des données
- ☐ Problèmes liés au caractère confidentiel / privé / sensible des données
- ☐ Problèmes liés à des dispositions législatives sectorielles
- ☐ Problèmes liés à une copropriété sur les données
- ☐ Problèmes liés à la préparation des données pour leur archivage
- ☐ Problèmes liés à la réutilisation des données
- ☐ Autres : .....

Phase 4 : Conditions de réutilisation des données  
11h40 – 12h20

Comment, en tant que réutilisateur potentiel des données, concevez-vous les conditions de leur accès auprès du système d'archivage ?

*Nous vous proposons différents scénarios, non limitatifs.*

*Vous pouvez cocher ceux que vous souhaitez.*

*Plusieurs réponses peuvent être possibles...*

## I. Utilisateurs potentiels des données

- ☐ Les chercheurs
- ☐ Les administrations de la recherche
- ☐ Le monde politique et les administrations publiques
- ☐ Entreprises
- ☐ Presse et médias
- ☐ Etudiants
- ☐ Toute autre personne
- ☐ .....
- ☐ .....

## II. Accès aux données

### Situation

- ☐ Portail accessible sur Internet
- ☐ Portail accessible sur Intranet
- ☐ Portail accessible des deux manières

### Accès au contenu

- ☐ Accès directement aux données (virtual library)
- ☐ Système de liens vers des données existant sur d'autres sites (portail d'information)
- ☐ Système de référencement de données (référothèque)
- ☐ .....
- ☐ .....

### Moyen d'accès aux données

- ☐ Accès libre aux données
- ☐ Accès aux données avec identification, login et mot de passe
- ☐ Accès aux données sur base d'un formulaire à remplir
- ☐ .....
- ☐ .....

### Dans le cas d'un accès direct aux données

- ☐ Accès nécessitant de s'enregistrer pour accéder à l'ensemble du portail d'archivage
- ☐ Accès ouvert uniquement aux fiches documentaires des ressources et nécessité de s'enregistrer dans le système pour avoir accès aux données
- ☐ Accès ouvert pour les fiches documentaires et les documents travaillés, mais fermé pour les données brutes
- ☐ Accès totalement ouvert pour l'ensemble du portail d'archivage
- ☐ Fixation de l'accès aux données par le dépositaire du document dans le portail selon des catégories prédéfinies : ex : données accessibles à tout le monde, données uniquement réservées à l'enseignement et à la recherche, ...
- ☐ .....
- ☐ .....

Accès payant ou gratuit

- ☐ Accès gratuit aux données
- ☐ Accès payant aux données
- ☐ Formule combinée
- ☐ .....
- ☐ .....

IV. Conditions susceptibles d'être imposées pour la réutilisation des données

- ☐ Obligation de citation de la source
- ☐ Interdiction du transfert de données à des tiers
- ☐ Obligation de mettre les données dérivées à la disposition du service d'archivage
- ☐ Obligation d'effacer les données dès la recherche terminée
- ☐ Interdiction d'utilisation dans certains buts (but commercial,...)
- ☐ Participation financière au fonctionnement du service
- ☐ .....
- ☐ .....

## ANNEXE 7. Exemples de contrats

---

### Chapitre 1. Exemples de licences de dépôt de données au centre d'archivage

---

#### 1. DANS <sup>1</sup>

##### Licentie overeenkomst

archiefnummer:  
deponeringsdatum:

Licentie overeenkomst betreffende [specificatie dataset]  
Titel van het onderzoek:

Bij deze licentie overeenkomst zijn de volgende partijen betrokken:

1. de organisatie of de persoon die gerechtigd is de digitale dataset(s) over te dragen en in beheer te geven, hierna te noemen: *de depotgever*
2. de organisatie die gerechtigd is de digitale dataset(s) te archiveren en te beheren, hierna te noemen: *de depotnemer*

De depotgever is:

Naam:  
Organisatie2:  
Postadres:  
Postcode:  
Plaats:  
Land:  
Telefoon:  
E-mail:

De depotnemer is:

Organisatie: DANS, Data Archiving and Networked Services, namens de KNAW (Koninklijke Nederlandse Akademie van Wetenschappen)  
Vertegenwoordigd door: Dr. P.K. Doorn, directeur  
Postadres: Postbus 93067  
Postcode: 2509 AB  
Plaats: Den Haag  
Land: Nederland  
Telefoon: (+31) (0)70 349 44 50  
E-mail: info@dans.knaw.nl

DANS is een KNAW-instituut mede ondersteund door NWO (Nederlandse Organisatie voor Wetenschappelijk Onderzoek) en is gevestigd te Den Haag.

---

<sup>1</sup> Disponible sur [http://www.dans.knaw.nl/nl/data\\_deponeren/licentiecontract\\_dans/](http://www.dans.knaw.nl/nl/data_deponeren/licentiecontract_dans/)

<sup>2</sup> Indien niet verbonden aan organisatie of instelling: s.v.p. invullen privé

De volgende bepalingen zijn op deze licentie overeenkomst van toepassing:

1. De licentie
  - a. De depotgever verleent de depotnemer een niet-exclusieve licentie om digitale data c.q. digitale dataset(s), zoals nader gespecificeerd in deze overeenkomst, hierna te noemen 'dataset' op te nemen in het data-archief van de depotnemer, door het overbrengen van de inhoud van de dataset op een, aan depotnemer ter beschikking staande drager, ongeacht op welke wijze en in welke vorm.
  - b. De depotnemer is gerechtigd, zulks met in achtneming van de overeengekomen toegangscategorie(en), (substantiële delen van) de dataset aan derden ter beschikking te stellen, door verspreiding, on line transmissie of transmissie in een andere vorm. Depotnemer krijgt daarbij het recht om, al dan niet in opdracht van derden, een kopie van de dataset te maken, dan wel derden toestemming te verlenen een kopie te downloaden.
2. De depotgever
  - a. De depotgever verklaart (enige) rechthebbende te zijn tot de dataset uit hoofde van de Databankenwet en (eventueel) de Auteurswet of anderszins en/of in dezen te handelen met de toestemming van de (mede) rechthebbende(n).
  - b. De depotgever vrijwaart de depotnemer van alle aanspraken die anderen jegens depotnemer zouden kunnen doen gelden ter zake van (de overdracht van) de dataset, de vorm en/of de inhoud ervan.
3. De depotnemer
  - a. De depotnemer zal, naar beste vermogen en middelen, de gedeponeerde dataset duurzaam archiveren, leesbaar en toegankelijk houden.
  - b. De depotnemer zal de dataset zoveel mogelijk ongewijzigd en in het oorspronkelijke software formaat houden voor zover dat mogelijk is, rekening houdend met de stand van de techniek en de kosten voor tenuitvoerlegging. De depotnemer heeft het recht om de vormgeving en/of functionaliteit van de dataset te wijzigen als dit noodzakelijk mocht zijn met het oog op de digitale duurzaamheid, verspreiding of hergebruik van de dataset.
  - c. Indien voor de toegangscategorieën "restricted access" "other access", zijnde respectievelijk de bijzondere gebruiksbepalingen A II, A III zoals aan het einde van deze overeenkomst is vermeld, gekozen is, zal de depotnemer, naar beste vermogen en middelen, zorg dragen voor doeltreffende (technische) voorzieningen teneinde te voorkomen dat onbevoegde derden substantiële delen van de dataset zouden kunnen opvragen en/of hergebruiken.
4. De dataset
  - a. De dataset waarop de licentie betrekking heeft, is in de bijlage bij deze overeenkomst nader gespecificeerd. Deze bijlage maakt integraal deel uit van deze overeenkomst.
  - b. De depotgever verklaart dat de dataset overeenstemt met de daarvoor afgegeven specificatie.
  - c. De depotgever verklaart dat de dataset geen gegevens of andere elementen bevat die in strijd zijn met de Nederlandse wet.
  - d. De depotgever vrijwaart de depotnemer tegen alle aanspraken van anderen betreffende de (inhoud van de) dataset.
  - e. De depotgever biedt de dataset aan op een wijze en via een medium dat door de depotnemer als geschikt en rendabel wordt geacht.
5. Verwijderen bestand/wijzigen toegankelijkheid
  - a. Indien daartoe zwaarwegende redenen bestaan heeft de depotgever het recht om depotnemer te verzoeken, de dataset –eventueel tijdelijk- niet ter beschikking te stellen aan derden. De depotnemer zal in dat geval de dataset wel in het data-archief behouden, maar vanaf dat moment derden niet langer toegang verschaffen tot (substantiële delen van) de dataset..
  - b. Indien daartoe zwaarwegende redenen bestaan heeft depotnemer het recht om de dataset geheel of gedeeltelijk uit het data-archief te verwijderen, dan wel de toegang daartoe tijdelijk of permanent te beperken of uit te sluiten. De depotnemer zal depotgever hiervan op de hoogte stellen.

6. Beschikbaarstelling aan derden:
  - a. De depotnemer zal de dataset beschikbaar stellen aan derden, overeenkomstig de met depotgever overeengekomen toegangscategorie(en): (AI) "Open access", (AII) "Restricted access", (AIII) "Other access".
  - b. De depotnemer zal de dataset uitsluitend beschikbaar stellen aan derden, met wie zij gebruiksbepalingen is overeengekomen. Tenzij met de depotgever anders is overeengekomen, gelden de algemene gebruiksbepalingen, zoals op de website van DANS (DANS EASY) zijn vermeld.
  - c. Indien de toegangscategorieën (AII) "restricted access" en/of (AIII) "other access" zijn overeengekomen, zal de depotnemer de dataset uitsluitend beschikbaar stellen aan bepaalde personen en/of organisaties, die daartoe zijn aangewezen door de depotgever.
  - d. Indien een dataset waarvoor een toegangscategorie (AIII) "other access" geldt, blijkens de door depotgever opgestelde specificatie- persoonsgegevens in de zin van de Wet Bescherming Persoonsgegevens WBP (Wet Bescherming Persoonsgegevens, wet van 6 juli 2000, Staatsblad 302, artikel 9 lid 3 en artikel 23 lid 2) bevat, zal depotnemer de dataset uitsluitend ter beschikking stellen indien dit wettelijk is toegestaan, waaronder in ieder geval dient te worden begrepen het ter beschikking stellen ten behoeve van wetenschappelijk, statistisch of historisch onderzoek.
  - e. In afwijking van het hiervoor bepaalde kan de depotnemer (substantiële delen van) de dataset beschikbaar stellen aan derden:
    - indien de depotnemer hiertoe wordt verplicht uit hoofde van wet- en regelgeving, een gerechtelijke uitspraak, of door een (toezichthoudende) instantie;
    - indien dit noodzakelijk is voor het behoud van de dataset c.q. het data-archief;
    - indien dit noodzakelijk is voor het (in opdracht c.q. ten behoeve van de depotnemer) verrichten van werkzaamheden aan het data-archief, voor de duur van deze werkzaamheden; (aan een soortgelijke instelling) indien depotnemer ophoudt te bestaan en/of haar activiteiten op het gebied van data-archivering beëindigt.
  - f. De depotnemer zal, op basis van de door de depotgever verstrekte documentatie bij de dataset, de metadata vrij beschikbaar stellen en publiceren. Onder metadata wordt verstaan: de beschrijvende informatie met betrekking tot de digitale bestanden. Overige, door de depotgever verstrekte, documentatie bij de dataset wordt vrij beschikbaar gesteld en gepubliceerd, tenzij de depotgever gespecificeerd aangeeft welke documenten niet vrij beschikbaar kunnen worden gesteld.
  - g. De algemene informatie over het onderzoek en de bij de dataset behorende metadata worden opgenomen in de databases en publicaties van de depotnemer, welke voor iedereen toegankelijk zijn.
7. Bepalingen met betrekking tot het gebruik door derden
  - a. De depotnemer verplicht derden, aan wie zij (substantiële delen van) de dataset ter beschikking stelt, het onderstaande te vermelden in eigen publicaties, indien daarbij uit de dataset afkomstige data gebruikt zijn:
    - De herkomst van de data (de gebruikte data zijn afkomstig uit een bij de depotnemer gedeponeerde bestand) en de identifier waaronder het databestand is opgeslagen;
    - Naam of namen van degene(n), die de dataset gecreëerd heeft/hebben;  
(indien afwijkend) Naam of namen van degene(n) die uit hoofde van de Databankenwet of de Auteurswet rechthebbende is/zijn tot de dataset;
    - De titel van de dataset.
    - Datum waarop de dataset gecreëerd is;
  - b. De depotnemer verplicht degene(n), aan wie zij de dataset ter beschikking stelt, deze dataset, indien hij/zij dit heeft/hebben aangevuld, dan wel heeft/hebben gewijzigd, zodanig dat er sprake is van een nieuwe dataset, ter zake van deze nieuwe dataset een (niet-exclusieve) licentie te verlenen aan depotnemer.
8. Overlijden depotgever  
Na het overlijden van de depotgever, dan wel na opheffing of beëindiging van diens organisatie, vervalt een eventueel voor de dataset geldende bijzondere gebruiksbepaling (A II) "Restricted access" en is automatisch de bijzondere gebruiksbepaling (A I) "Open access" van kracht.
9. Aansprakelijkheid

- a. De depotnemer is niet aansprakelijk voor het geheel of gedeeltelijk verloren gaan van de dataset.
  - b. De depotnemer is niet aansprakelijk voor eventuele schade als gevolg van een handelen of nalaten door derden, aan wie de depotnemer de dataset ter beschikking heeft gesteld.
10. Duur van de overeenkomst, beëindiging
- a. Deze overeenkomst gaat in op de hieronder vermelde datum van ondertekening door de depotnemer, hierna te noemen de deponeringsdatum en eindigt van rechtswege na verloop van 16 jaar ingevolge de Databankenwet. Na het verstrijken van voormelde termijn behoudt de depotnemer de dataset in het data-archief en blijft de depotnemer gerechtigd de dataset ter beschikking te stellen van derden. De licentieovereenkomst zal na afloop van een periode van 16 jaar worden voortgezet indien en voor zover dit noodzakelijk mocht zijn om eventuele rechten uit hoofde van de Auteurswet te respecteren.
  - b. In afwijking van het bepaalde onder a. eindigt deze overeenkomst zodra de dataset permanent uit het data-archief verwijderd is op grond van artikel 5 van deze overeenkomst.
  - c. Indien de depotnemer ophoudt te bestaan of haar activiteiten op data-archiveringsterrein beëindigt zal de depotnemer trachten de bestanden bij een gelijkwaardige instelling onder te brengen, welke de overeenkomst, indien mogelijk, met depotgever voortzet.
11. Procesbevoegdheid
- DANS is bevoegd, maar niet verplicht, om zelfstandig op te treden tegen inbreuken op het Auteursrecht en/of een ander intellectueel eigendomsrecht van de rechthebbende(n) op de dataset en/of op de daarin opgenomen data.
12. Toepasselijk recht
- Op deze overeenkomst is het Nederlands recht van toepassing.
- 

Bijzondere bepalingen met betrekking tot het gebruik:

De volgende bijzondere bepalingen met betrekking tot het gebruik van de dataset door derden zijn van toepassing op deze licentie overeenkomst.

- A. De depotgever dient bij de bijzondere bepaling A tenminste één van de drie toegangs-categorieën I, II, III of IV aan te kruisen. Indien géén vak is aangekruist geldt standaard het onder I vermelde,

De dataset mag door de depotnemer worden gedistribueerd en ter beschikking worden gesteld op de hierna onder I, II, III of IV door de depotgever aangegeven wijze, zulks mede met inachtneming van het in artikel 6 van de overeenkomst gestelde met betrekking tot persoonsgegevens:

**[Open access: onbeperkte toegang]**

☐ I. Aan iedereen. De dataset mag door de depotnemer met onmiddellijke ingang vanaf de deponeringsdatum aan iedereen worden gedistribueerd en ter beschikking worden gesteld.

**[Restricted access: beperkte toegang]**

☐ II. Uitsluitend aan personen en organisaties, die daartoe zijn aangewezen door de depotgever.

Deze bepaling geldt voor een periode van ..... maanden<sup>3</sup> te rekenen vanaf de deponeringsdatum met een maximum van 24 maanden. Na het verstrijken van deze periode wordt automatisch de bijzondere gebruiksbepaling I (Open Access) van kracht, tenzij de depotgever heeft aangegeven de bijzondere gebruiksbepaling II (Restricted Access) te willen verlengen. Dit laatste is mogelijk met steeds een eenzelfde periode van maximaal 24 maanden.

**[Other access: geen toegang via de website van DANS]**

---

<sup>3</sup> aantal maanden invullen, maximaal 24.



☐ III. Door DANS wordt op geen enkele wijze via de website van DANS toegang tot de dataset verleend. Toegang wordt eventueel via een andere, nader met DANS overeen te komen, wijze overeengekomen

- B. Indien meerdere bestanden ineens worden overgedragen, en voor één of enkele bestanden afwijkende toegangsbepalingen van kracht zijn, moeten deze in de bestandenlijst (bijlage) worden vermeld.

## 2. UKDA <sup>4</sup>

### Licence Agreement

This Agreement is made between the UK Data Archive (a department of the University of Essex and not a separate legal entity) of Wivenhoe Park, Colchester, C04 3SQ (the "University") and the Depositor (as defined below).

#### GUIDANCE NOTES

The University through the UK Data Archive acquires, preserves and disseminates data collections. The Depositor owns the data collection as described below and wishes to license these materials to the University.

This is a non-exclusive licence which ensures that copyright in the original data is not transferred by this Agreement and provides other safeguards for the Depositor, such as, requesting acknowledgement in any publications arising from future research using the data. It permits use of the Data Collection for the Purpose specified in Part C. Access to the Data Collection will only be available to registered users who have agreed to abide by the licence conditions.

#### LICENCE TERMS AND CONDITIONS

##### 1 Definitions and Interpretation

1.1 In this Agreement the following words have the following meanings:

'Agreement' - this Agreement including its schedules and all and any of the Data Collection Deposit Forms completed by the Depositor

'Authorised User' individuals registered by the University to use the Data Collection, or a member of an institution registered by the University to use the Data Collection under a site licence or re-distribution agreement, or individuals who are not formally registered by the University to use the Data Collection but who have signed an access agreement in relation to work being undertaken by a registered user (e.g. students undertaking courserelated work who have signed an Access Agreement for Teaching [Academic Sector] form, which has been accepted by the University)

'Commercial Purposes' use of the Data Collection either directly or indirectly for any Commercial Purpose (whether for gain or not)

'Data Collection' the material to be provided by the Depositor under the title in the Data Collection Deposit Forms under the terms and conditions of this Agreement. The terms "dataset" or "study" may be used in some Archive documentation to mean Data Collection

---

<sup>4</sup> Disponible à l'adresse <http://www.esds.ac.uk/aandp/create/licence.asp>

'Data Collection Deposit Forms ' means the set of forms made available on the UKDA web site that have been completed by the Depositor describing the contents of the Data Collection that have been submitted to, and accented by, the University as from time to time

'Depositor' means the person or persons set out in page 1 of this Agreement

'Educational and Research Purposes' use of the Data Collection for education, private study or private research provided that such is not for Commercial Purposes

'Registered User' individuals registered by the University to use the Data Collection or a member of an institution registered by the University to use the Data Collection under a site licence or re-distribution agreement

'Term' this Agreement shall take effect on the date when executed by the last party and shall continue for the duration of copyright in the Data Collection unless terminated in accordance with Clause 7 of this Agreement.

'UK Data Archive Holdings' means all the data collections held by the University Created 1

## 2 Licence

2.1 The Depositor hereby grants a worldwide, royalty-free licence of the Data Collection to the University for the Term for the following purposes, without limitation:

- 2.1.1 distribute copies of the Data Collection to Registered Users in a variety of media formats;
- 2.1.2 promote and advertise the Data Collection in any publicity (in any form) for the University;
- 2.1.3 to catalogue, enhance, validate and document the Data Collection;
- 2.1.4 to store, translate, copy or re-format the Data Collection in any way to ensure its future preservation and accessibility;
- 2.1.5 incorporate metadata or documentation in the Data Collection into public access catalogues for the Data Collections (together referred to as "the Purpose").

## 3 Depositor's Rights and Undertaking

3.1 The Depositor is free to use or publish the Data Collection elsewhere.

3.2 The Depositor does not warrant or guarantee the Data Collection in terms of the comprehensiveness, accuracy, reliability, or otherwise of its contents.

3.3 The Depositor hereby warrants and undertakes as follows:

- 3.3.1 that it is entitled to grant the rights contained in this Agreement to the University and the exercise by the University of the rights assigned to it under this Agreement will not infringe the rights of any third party;
- 3.3.2 that the Data Collection does not and will not contravene any laws as in force from time to time, including but not limited to the law relating to defamation, or obscenity;
- 3.3.3 that the Depositor is not under any obligation or restriction created by law, contract or otherwise which would in any manner or to any extent prevent or restrict the Depositor from entering into and fully performing this Agreement;
- 3.3.4 to promptly notify the University of any change of copyright ownership affecting the Data Collection;
- 3.3.5 to promptly notify the University of any confidentiality, privacy or data protection issues pertaining to the Data Collection.

3.4 The Depositor hereby indemnifies and will keep indemnified the University and its employees and appointed agents against all damages, losses, claims, costs and expense for which they become legally liable to the extent that this shall arise out of any negligent act or omission or breach of this Agreement committed by or on behalf of the Depositor. This indemnity shall survive the termination of this Agreement for any reason.

## 4 The University's Rights and Responsibilities

4.1 The University shall:

4.1.1 take reasonable measures to prevent unauthorised access to duplication of or distribution of the Data Collection whilst it is in the University's possession or under its control;

4.1.2 permit Registered Users to access and use the Data Collection, or any part of it;

4.1.3 draw the following information to the attention of each Registered User as part of the registration process or as part of the data delivery process (subject that such information may be changed from time to time):

No part of these materials may be reproduced, stored in, or introduced into a retrieval system, or transmitted, in any form, or by any means (electronic, mechanical, photocopying, recording or otherwise) without the prior written permission of the UK Data Archive.

4.1.4 request Registered Users publishing any work based in whole or in part on the Data Collection to acknowledge the original data creators, depositors or copyright holders, the funders of the Data Collections (if different) and the UK Data Archive, and to acknowledge Crown Copyright where appropriate. Any publication, whether printed, electronic or broadcast, based wholly or in part on these materials should carry a statement that the original data creators, depositors or copyright holders, the funders of the Data Collections (if different) and the UK Data Archive bear no responsibility for their further analysis or interpretation;

4.1.5 not be under any obligation to take legal action on behalf of the Depositor or other rights-holders in the event of breach of any intellectual property rights or any other right in the Data Collection deposited;

4.1.6 not be under any obligation to reproduce, transmit, broadcast, or display the Data Collection in the same formats or resolutions as those noted in the Data Collection Deposit Forms;

4.1.7 not be under any obligation to include the Data Collection or any part of it within the UK Data Archive's Holdings and retain the right to remove all or any part of the Data Collection from the UK Data Archive's Holdings.

4.2 While reasonable care will be taken to preserve the physical integrity of the Data Collection, the University shall incur no liability, either express or implicit, for the Data Collection or for the loss of or damage to any of the Data Collection.

4.3 The copyright in any additional data added by the University to the Data Collection, and any search software, user guides, documentation and any other intellectual property rights that are prepared by the University to assist users in using the Data Collection will belong to the University and any other parties that the University may choose to enter into an agreement with to produce such materials.

## 5 Royalties

5.1 No royalties shall be paid for the use of the Data Collection for Educational Purposes, archiving, or publicity for the Purpose set out in Clause 2 above.

5.2 For Commercial Purposes, royalty payments may be collected by the University on behalf of the Depositor, subject to a further written agreement and the University shall not have any duty to pay such royalties unless it executes such a further written agreement.

## 6 Notices

6.1 All notice under this Agreement shall be in writing and shall be sent to the address of the recipient set out in this Agreement or to such other address as the recipient may have notified from time to time. Any notice may be delivered personally or by first class post or by fax or by email and shall be deemed to have been served if by hand when delivered, if by first class post 48 hours after posting, if by fax when confirmation of transmission is received and if by email, when confirmation of receipt is received from the system of the recipient. If no reply is received to a notice under this Agreement the consent of the recipient will be deemed to have been given after thirty (30) days have elapsed from the issue of that notice.

## 7 Termination

7.1 In addition to any other remedy, the University on the one hand and the Depositor on the other may terminate this Agreement immediately in the event of any breach of this Agreement which

cannot be remedied or is not remedied within thirty (30) days of the party in breach being requested to do so by the other party and all obligations shall cease except where specified otherwise under this Agreement.

7.2 Where there is no breach, either party may terminate this Agreement upon six months written notice and if the Depositor terminates the Agreement under this Clause 7.2 the University shall be entitled to charge the Depositor for such reasonable costs as have been incurred in archiving and cataloguing, and any other investment of resources in the Data Collection, prior to its withdrawal.

## 8 General

8.1 This Agreement is binding on and will benefit the successors and assigns of the parties.

8.2 The University will not assign, transfer or subcontract the Agreement or any rights under it without prior written consent of the Depositor.

8.3 This Agreement constitutes the entire agreement between the parties. No variation will be effective unless agreed in writing by the University.

8.4 If any part of this Agreement is held unlawful or unenforceable that part shall be struck out and the remainder of this Agreement shall remain in effect.

8.5 This Agreement does not create any partnership or joint venture between the parties.

8.6 No delay, neglect, or forbearance by either party in enforcing its rights under this Agreement shall be a waiver of or prejudice of those rights.

8.7 This Agreement is governed and interpreted in all respects under the laws of England and Wales and shall be subjected to the jurisdiction of the courts of England and Wales.

8.8 The Depositor and the University shall be under no liability for any loss or for any failure to perform any obligation hereunder due to causes beyond their control, including but not limited to industrial disputes of whatever nature, Acts of God, hostilities, force majeure or any circumstances which they could not reasonably foresee and provide against.

---

## Chapitre 2. Exemples de contrats de réutilisation des données

---

### 1. DANS

#### Gebruiksbepalings DANS <sup>5</sup>

Degene aan wie door DANS data ter beschikking worden gesteld, hierna te noemen de data-gebruiker, verklaart zich accoord met de hiernavolgende algemene gebruiks-bepalingen. De data-gebruiker mag uitsluitend onder deze voorwaarden gebruikmaken van de data.

#### ALGEMENE GEBRUIKSBEPALINGEN DANS

##### 1. Bibliografische referentie

De datagebruiker zal altijd in publicaties, waarbij gebruik is gemaakt van data, die afkomstig van een van de DANS-datasets, vermelden:

- De herkomst van de data (de gebruikte data zijn afkomstig uit een bij DANS gedeponeerde dataset) en de identifier waaronder de datasets is opgeslagen (= ID-nummer zoals vermeld in de documentatie van de dataset);
- Naam of namen van degene(n), die de dataset gecreëerd heeft/hebben, doch in ieder geval de naam of namen van degene(n) die uit hoofde van de Databankenwet of de Auteurswet rechthebbende is/zijn tot de dataset (dit is de naam die vermeld staat bij "creator" in de documentatie van de dataset);
- De titel van de dataset;
- De datum van creatie van de dataset (= verzamelperiode, zoals vermeld in de documentatie van de dataset);
- 

##### 2. Publicaties

De data-gebruiker zal DANS de bibliografische gegevens (waaronder in ieder geval standaard identificatienummers als ISBN, ISSN of DOI, indien aanwezig) toekomen van iedere gedrukte of digitale publicatie die niet via Internet beschikbaar is en waarvoor gebruik is gemaakt van data die afkomstig zijn van één van de DANS-datasets. Indien de publicatie via Internet beschikbaar is, zal de data-gebruiker de URL daarvan doorgeven aan DANS.

##### 3. Nieuwe versie

De data-gebruiker verplicht zich, indien hij of zij een dataset zodanig aanvult of wijzigt dat er sprake is van een nieuwe versie of een nieuwe dataset, deze nieuwe versie c.q. nieuwe dataset bij DANS te deponeren door het te deponeren in DANS EASY.

##### 4. Weergeven persoonsgegevens

De data-gebruiker stemt ermee in dat zijn of haar persoonsgegevens gebruikt zullen worden voor intern DANS gebruikersonderzoek. De data-gebruiker stemt er bovendien mee in dat de door hem of haar bij de registratie verstrekte gegevens weergegeven zullen worden op het voor geregistreerde gebruikers toegankelijke deel van de DANS website bij elk dataset die door de data-gebruiker naar de eigen computer wordt gekopieerd ("gedownload"). Het gaat hierbij om de velden naam, organisatie en functie.

##### 5. Verspreiding van de dataset

De dataset mag uitsluitend voor persoonlijk gebruik worden gekopieerd. De dataset mag niet voor commerciële doeleinden worden gebruikt of worden (door)verkocht.

##### 6. Bescherming persoonsgegevens

Datasets die persoonsgegevens, in de zin van de Wet Bescherming Persoonsdata bevatten, mogen uitsluitend ten behoeve van wetenschappelijk, statistisch of historisch onderzoek gebruikt worden.

---

<sup>5</sup> Disponible sur le site <http://www.dans.knaw.nl>

Voor gebruikers van datasets met persoonsgegevens stelt DANS een werkwijze volgens de *Gedragscodex voor gebruik van persoonsgegevens in wetenschappelijk onderzoek* VSNU verplicht. Op de data-gebruiker rust de plicht altijd de vertrouwelijkheid van persoonsgegevens te waarborgen.

#### 7. Auteursrechten van derden

Indien op een dataset of delen daarvan auteursrechten of andere exclusieve rechten rusten van anderen dan degene, die de dataset gedeponeerd heeft (depotgever), zal de data-gebruiker deze dataset, of delen daarvan, niet zonder toestemming van DANS en/of de depotgever openbaar maken. .

#### 8. Aansprakelijkheid data

DANS is op geen enkele wijze aansprakelijk voor fouten of onjuistheden in de verstrekte gegevens of de bijbehorende documentatie. De data-gebruiker wordt verzocht eventuele aangetroffen onjuistheden aan DANS door te geven.

#### 9. Aansprakelijkheid publicaties

DANS is niet aansprakelijk of verantwoordelijk voor de inhoud van publicaties waarvoor gebruik is gemaakt van data die afkomstig zijn van een van de DANS-datasets.

## 2. UKDA <sup>6</sup>

### END USER LICENCE

This Agreement is made between you and the University of Essex (also referred to as the "registrar") and the service funders in order to provide you (the "End User") with the right to use the collections provided via the ESRC/JISC Census Programme, the Economic and Social Data Service (ESDS) and the UK Data Archive, according to the terms below.

In this agreement:

"Data Team" means in relation to a particular data collection, the registrar, the relevant data service providers, and (to the extent that the Special Conditions and/or metadata specific to a particular data collection expressly provide) the service funders, data collection funders and/or original data creators or depositors.

"data service provider" means the persons or organisations that directly provide you with the data collections (on behalf of the service funder). The data service provider for a particular data collection is identified in the Special Conditions and/or metadata applicable to that data collection;

"service funder" means the persons or organisations that fund the data service provider as defined above. The service funder for a particular data collection is identified in the Special Conditions and/or metadata applicable to that data collection;

"data collection funder" means the persons or organisations that funded the collection and/or creation of the data collections. The data collection funder for a particular data collection is identified in the Special Conditions and/or metadata applicable to that data collection;

"original data creator or depositor" means the persons or organisations that originally collected, created or deposited the materials making up the data collections and/or who own the intellectual property rights in the data collections. The original data creator or depositor for a particular data collection is identified in the Special Conditions and/or metadata applicable to that data collection;

---

<sup>6</sup> Disponible à l'adresse <http://www.data-archive.ac.uk/aandp/access/licence.asp>

"registrar" means the person or organisation responsible for the system that registers End Users and issues them with End User Licences (being the University of Essex);

"Special Conditions" means any further conditions applicable to the use of one or more data collections by an End User, as notified to the End User in accordance with paragraph 5 of the End User Licence.

"metadata" means any additional or bibliographic information about one or more of the data collections, as notified to the End User from time to time. Metadata may be supplied by electronic means.

I (the "End User") agree to the following conditions of use in consideration of the data collections being made available to me through the various contributions of each member of the Data Team:

1. To use the data collections only in accordance with this End User Licence and to notify promptly the registrar and the data service provider of any breach of its terms in writing or of any infringements of the data collections of which I become aware.
2. To use and to make personal copies of any part of the data collections only for the purposes of not-for-profit research or teaching or personal educational development. To obtain permission prior to using part or all of the data collections for commercial purposes by contacting the registrar and/or relevant data service provider, where relevant, in order to obtain an appropriate licence from the rights holder(s) in question or their permitted licensee if one is available.
3. That this Licence does not operate to transfer any interest in intellectual property from the data collection funders, service funder(s), the data service providers, the original data creators, producers, depositors, copyright or other right holders (including without limitation the ONS or the Crown ) to me. That any rights subsisting in materials derived now or in the future from the data collections which are the intellectual property of the Crown are hereby assigned (by way of assignment of present and future intellectual property) to the Crown by this Licence to the extent not already vested in the Crown. To take all steps necessary to give effect to this Clause (including by executing further written documentation).
4. That the Licence and the data collections are provided by the Data Team on an "as is" basis and without warranty or liability of any kind. Any representations or warranties given by any member of the Data Team relating to this licence, expressed or implied, are excluded to the maximum extent permitted by law.
5. To abide by any further conditions notified to me from time to time by the registrar or the relevant data service provider that may apply to the access to, or use of, specific materials within the data collections or particular data collections. Notice of further conditions under this paragraph may be given to me by electronic means, for example, by way of a pop-up window upon my ordering one or more data collections. My acceptance of the further conditions shall be required before I gain access to the data collections in question. In this Agreement such further conditions are referred to as Special Conditions.
6. To give access to the data collections, in whole or in part, or any material derived from the data collections, only to registered End Users who have entered into an End User Licence and accepted the relevant Special Conditions necessary to access and use the data collections (with the exception of data collections or material derived from data collections supplied for the stated purpose of teaching or included in publications made for the purposes set out in paragraph 2).
7. To ensure that the means of access to the data (such as passwords) are kept secure and not disclosed to a third party except by special written permission or licence obtained from the original data service provider.
8. To preserve at all times the confidentiality of information pertaining to individuals and/or households in the data collections where the information is not in the public domain. Not to

use the data to attempt to obtain or derive information relating specifically to an identifiable individual or household, nor to claim to have obtained or derived such information. In addition, to preserve the confidentiality of information about, or supplied by, organisations recorded in the data collections. This includes the use or attempt to use the data collections to compromise or otherwise infringe the confidentiality of individuals, households or organisations.

9. To acknowledge, in any publication, whether printed, electronic or broadcast, based wholly or in part on the data collections, the original data creators, depositors or copyright holders, the service funders and the data service provider(s) in the form specified on the data distribution notes or in accompanying metadata received with the dataset or notified to me and without prejudice to paragraph 5 above to comply with any restrictions on my use of the data collections referred to or referenced therein or otherwise notified to me from time to time. To cite, in any publication, whether printed, electronic or broadcast, based wholly or in part on the data collections, the data collections used in the form specified on the data distribution notes or in accompanying metadata received with the dataset or notified to me.
10. To supply the relevant data service provider with the bibliographic details of any published work based wholly or in part on the data collections.
11. That the members of the Data Team may hold and process any personal data submitted by me for validation and statistical purposes, and for the purposes of the management of the service or for any other lawful purpose notified to me and to which I have consented under this Agreement in relation to a particular data collection, and they may also pass the information on to other parties such as: (i) depositors and distributors of material contained in or accessed via the data service provider; (ii) copyright and other intellectual property rights owners whose material is held by the data service provider; as well as (iii) each member of the Data Team's organisation and (iv) my own institution or organisation, in compliance with the Data Protection Act 1998.
12. To notify the data service provider of any errors discovered in the data collections.
13. That any personal data submitted by me is accurate to the best of my knowledge, and that any changes in that personal data, including my educational or employment status, will be made known to the registrar at the earliest possible opportunity.
14. To meet any charges that may from time to time be levied by any member of the Data Team for the supply of the data collections including, where relevant, annual service fees and royalty fees.
15. At the conclusion of my research (or if earlier at any time at the request of a member of the Data Team), to offer for deposit in the data collection(s) on a suitable medium and at my own expense any new data collections which have been derived from the materials supplied or which have been created by the combination of the data supplied with other data. The deposit of the derived data collection(s) will include sufficient explanatory documentation to enable the new data collection(s) to be accessible to others.
16. I understand that breach of any of the provisions of this Agreement will lead to immediate termination of my access to all services provided by the Data Team either permanently or temporarily, at the discretion of a member of the Data Team, and may result in legal action being taken against me. I understand that where there is no breach of this Licence, it may be terminated, or its terms altered, by a member of the Data Team either after 30 days notice; or, if a service charge has been paid in advance, at the end of the period for which payment has been made, whichever is the longer. The failure to exercise or delay in exercising a right or remedy provided by this Agreement or by law does not constitute a waiver of the right or remedy or a waiver of other rights or remedies.

## DISCLAIMERS

To the extent that applicable law permits:



- a. The members of the Data Team bear no legal responsibility for the accuracy or comprehensiveness of the data supplied.
- b. The members of the Data Team accept no liability for, and I will not be entitled to claim against them in respect of, any direct, indirect, consequential or incidental damages or losses arising from use of the data collections, or from the unavailability of, or break in access to, the service, for whatever reason.
- c. Whilst steps have been taken to ensure all licences, authorisation and permissions required for the granting of this Licence have been obtained, this may not have been possible in all cases, and no warranties or assurance are given in this regard. To the extent that additional licences, authorisations and permissions are required to use the data collections in accordance with this Licence, it is the End User's responsibility to obtain them.
- d. I agree to indemnify and shall keep indemnified each member of the Data Team against any costs, actions, claims, demands, liabilities, expenses, damages or losses (including without limitation consequential losses and loss of profit, and all interest, penalties and legal and other professional costs and expenses) arising from or in connection with any third party claim made against any member of the Data Team relating to my use of the data collections or any other activities in relation to the data where such use is in breach of this licence.

If the whole or any part of a provision of this Agreement is void, unenforceable or illegal for any reason, that provision will be severed and the remainder of the provisions of this Agreement will continue in full force and effect as if this Agreement had been executed with the invalid provision eliminated.

This Agreement may be enforced separately in relation to each data collection provided to the End User by any member of the Data Team and the End User. No other persons may enforce this Agreement under the Contract (Rights of Third Parties) Act 1999.

This Agreement (which is the entire agreement between the parties and supersedes any previous agreement between them) may be varied in writing by agreement of the relevant service funders, the registrar, and the End User (who may give its consent to such variations by electronic means). No consent from any other party is required to vary or rescind this Agreement.

This Agreement and any documents to be entered into pursuant to it shall be governed by and construed in accordance with the laws of England and Wales and each Party irrevocably submits to the exclusive jurisdiction of the courts of England and Wales over any claim or matter arising under or in connection with this Agreement and the documents entered into pursuant to it.

### **3. ZA,<sup>7</sup>**

#### **Regulations of use**

##### **1. Introduction**

The Central Archive for Empirical Social Research (ZA) is a service institute for social sciences. Apart from further infrastructure tasks the ZA archives scientific data and document stocks, especially from survey research, and makes this material available for further use.

---

<sup>7</sup> Disponibles à l'adresse [http://www.gesis.org/en/data\\_service/order/usage\\_regulations.htm](http://www.gesis.org/en/data_service/order/usage_regulations.htm)

The archive holding includes the data depositor's original data and records as well as additional ZA material, which is the result of standardised documentation and processing (e.g. Codebook, adjusted data sets, etc.).

Data and documents are exclusively made available on the basis of these regulations for use.

In the context of the international archive convention, users living outside Germany are asked to direct their requests regarding data to their home archive first.

## 2. General access conditions

As far as not explicitly indicated differently, the ZA makes data and documents available only for scientific analysis carried out by academic research and teaching. Institutes and persons outside academic research and doctrine can apply for provision in written form.

The acquisition of data and documents is done within the limits of access categories (please see 3.). The respective data depositor determines access limitations.

Independent of these access limitations for the use of data and documents, everybody is allowed to look at the documents – as far as not decided differently.

## 3. Access categories

The provision of data and documents is regulated by the following access categories. They are noted in the ZA [Data Holding Catalogue](#), too.

Category 0 Data and documents are released for everybody.

Category A Data and documents are released for academic research and teaching.

Category B Data and documents are released for academic research and teaching, if the results won't be published.  
If any publications, or any further work on the results is planned, a permission must be obtained by ZA.

Category C Data and documents are only released for academic research and teaching after the data depositor's written authorisation. For this purpose the ZA obtains a written permission with specification of the user and the analysis intention.

## 4. Request and provision of material

To [order](#) material from the archive holding, forms are provided in which all necessary information for the request processing can be entered. The material is only made available for the intention the ZA or the data supplier was notified about. The passing on of the material to a third party is not allowed.

## 5. Completion of the project

The user assumes responsibility to inform the ZA about the completion of the project the material has been used for. To prevent misuse the data has to be deleted or the data medium has to be made unusable after completion of the project. In case a further utilisation is intended, the user must apply to the ZA for new use permission.

6. Quotation liability, specimen copy

The user assumes responsibility to quote all used documents according to scientific conventions and to transfer two specimen copies of his/her publication to the Central Archive.

7. ZA scale of charges and fees

The ZA [scale of charges and fees](#) regulate the fees for the provision of data and documents from the ZA.

8. Final clause

Changes of the regulations for use and the scale of charges become effective from 1.1.2002.

## **ANNEXE 8. International workshop on the creation of a Belgian Social Data Archive – Programme (16 october 2007)**

---

### **International workshop on the creation of a Belgian Social Data Archive**

**DATE:** 16 October 2007  
**TIME:** 11:15 to 16:00  
**LOCATION:** Belgian Federal Science Policy Office  
Rue de la Science 8 Wetenschapsstraat  
B-1000 Brussels  
[http://www.belspo.be/belspo/about/acces\\_en.stm](http://www.belspo.be/belspo/about/acces_en.stm)

**TIMETABLE :**

11:15 to 11:30  
Welcome

11:30 to 12:00

***Key issues of the Belgian Social Data Archive (SODA) project***

Outline of the exploratory research regarding the opportunity to create a Belgian data archive for the social sciences and of its conclusions.

*By the research team*

12:00 to 12:30

***Social data archiving: the German experience***

Presentation of the Zentralarchiv für empirische Sozialforschung, one of the most developed European data archiving centres.

*By Ekkehard Mochmann and Meinhard Moschner*

12:30 to 12:45

Question-time

12:45 to 14:00

Lunch

14:00 to 14:30

***The Swedish experience on how to create a social data archive today***

Presentation of the challenges faced by the Swedish team when recently renewing their social data archive.

*By Magnus Stenbeck*

14:30 to 14:45

Question-time

14:45 to 15:30

***International cooperation and exchange of social data***

Presentation of Council of European Social Science Data Archives (CESSDA) and overview of the best European practices in social data archiving

*By Kevin Schürer, Director of the UK Data Archive and of CESSDA*

15:30 to 16:00

Questions and discussion.

16:00

End of the meeting

## **ANNEXE 9. Règlement général en matière de propriété, de protection et de valorisation des résultats des recherches réalisées au sein des FUNDP**

---

*(Décision du Conseil d'administration n° 523 du 28 juin 2004)*

### **Préambule**

Les FUNDP, dont la recherche constitue une des missions essentielles, souhaitent protéger et valoriser les résultats de leurs recherches conformément au présent règlement.

### **1. Définitions**

#### **1.1. Chercheur**

Au sens du présent règlement, on entend par Chercheur toute personne effectuant, seule ou en équipe, des recherches, études ou prestations au sein ou sous la responsabilité des FUNDP. Sont visés dans cette acception, tout membre du personnel académique, scientifique, administratif, technique et ouvrier, ainsi que les boursiers, étudiants, stagiaires et collaborateurs effectuant des recherches au sein des FUNDP ou sous la responsabilité de celles-ci.

#### **1.2. Résultats des recherches**

Au sens du présent règlement, on entend par Résultats des recherches, tout type de créations intellectuelles réalisées par le Chercheur ou de résultats obtenus par celui-ci, soit dans l'accomplissement de ses fonctions pour les FUNDP, soit par le recours à des techniques ou des moyens matériels ou humains mis à sa disposition par les FUNDP, ainsi que les droits de propriété intellectuelle ou industrielle y afférant.

Les Résultats des recherches comprennent notamment les inventions, les programmes d'ordinateur, les bases de données, les obtentions végétales, les informations scientifiques ou techniques non accessibles au public, les dessins et modèles et les topographies de produits semiconducteurs.

Pour l'application du présent règlement, sont cependant exclus des Résultats des recherches :

- les oeuvres protégées exclusivement par le droit d'auteur et qui ne sont pas des programmes d'ordinateur ;
- les inventions libres, à savoir les inventions réalisées par un Chercheur dans un domaine totalement extérieur à celui de ses recherches aux FUNDP et sans avoir recours à des techniques ou des moyens matériels ou humains, mis à sa disposition par les FUNDP.

### **2. Objet**

Le présent règlement a pour objet de fixer les règles applicables à la propriété, la protection et la valorisation des Résultats des recherches réalisées au sein des FUNDP.

### **3. Champ d'application**

Le présent règlement s'applique à tout Chercheur qui s'engage à en respecter les dispositions conformément aux termes de son contrat de travail, de son statut ou du document qui lui a été remis à cet effet.

### **4. Propriété des Résultats des recherches**

**4.1.** Les FUNDP sont propriétaires des Résultats des recherches obtenus par le Chercheur.

**4.2.** Le Chercheur conserve la possibilité de procéder à des publications ou communications scientifiques relatives aux Résultats des recherches conformément aux pratiques admises par la communauté scientifique et sous réserve des limites énoncées ci-après.

**4.3.** Dans le cas où un partenaire extérieur est impliqué dans le financement ou l'exécution de la recherche, la propriété des Résultats des recherches est régie par la convention liant les FUNDP au dit partenaire. Les droits et obligations du Chercheur et des FUNDP restent régis par le présent

règlement sous réserve des dispositions dérogatoires éventuellement contenues dans une telle convention.

## **5. Protection et valorisation des Résultats des recherches**

**5.1.** Le Chercheur s'engage à informer au plus tôt le promoteur de la recherche et la Cellule de Valorisation de l'Administration de la Recherche des FUNDP de tous Résultats des recherches qui pourraient justifier le dépôt d'un brevet ou faire l'objet d'une autre forme de valorisation, préalablement à toute publication ou autre communication au public. Le Chercheur remplira à cet effet le formulaire de déclaration n° ADRE 001 disponible auprès de la Cellule de Valorisation et annexé au présent Règlement.

Les FUNDP s'engagent à informer le Chercheur de leur décision quant aux mesures de protection et de valorisation qu'elles entendent attacher aux Résultats des recherches mentionnés dans cette déclaration et ce dans un délai maximum de trois mois. Aucune publication ou autre communication au public relatives aux Résultats des recherches du Chercheur n'aura lieu pendant cette période sauf accord explicite de la Cellule de Valorisation de l'Administration de la Recherche des FUNDP.

**5.2.** Les FUNDP décident librement et au cas par cas des mesures de protection et de valorisation qu'elles entendent attacher aux Résultats des recherches.

**5.3.** Si les FUNDP décident de protéger ou valoriser ces Résultats des recherches, le Chercheur apporte aux FUNDP toute l'assistance et la collaboration nécessaires à la protection et à la valorisation de ces Résultats des recherches. Notamment, il fournit les données scientifiques et techniques nécessaires à la rédaction du ou des brevet(s) et donne tout pouvoir aux FUNDP pour demander, partout dans le monde, les protections nécessaires. Il veillera également à ce que l'exercice du droit visé à l'article 4.2. ne porte pas préjudice à cette protection ou valorisation.

**5.4.** Si les Résultats des recherches donnent lieu au dépôt d'un brevet, il sera déposé au nom des FUNDP ou de toute autre personne physique ou morale qu'elles indiquent. Le brevet désignera comme inventeur chaque Chercheur ayant contribué de manière significative à l'invention décrite dans le brevet, conformément aux dispositions légales et réglementaires applicables. Il est loisible au Chercheur de notifier à la Cellule de Valorisation son refus d'être désigné comme inventeur.

**5.5.** Si les FUNDP décident de ne pas ou de ne plus protéger ou valoriser les Résultats des recherches et sous réserve des droits d'un partenaire extérieur, elles en informent le Chercheur concerné en temps utile. Avec son accord, les Résultats des recherches sont cédés au Chercheur par les FUNDP. Les FUNDP concluront un contrat à cet effet avec le Chercheur et conserveront en tous cas le droit d'utiliser les Résultats des recherches à des fins d'enseignement et de recherche.

## **6. Confidentialité**

**6.1.** Le Chercheur reconnaît qu'il pourrait avoir accès, dans le cadre de ses activités au sein des FUNDP, à des informations scientifiques et techniques non accessibles au public. Sans préjudice d'une convention spécifique, il s'engage à respecter la confidentialité de celles-ci et en conséquence, à ne pas les divulguer directement ou indirectement à des tiers pendant l'exercice de ses fonctions et Durant une période de deux ans après la cessation de celles-ci.

**6.2.** Par ailleurs, avant la cessation de ses fonctions de Chercheur, le Chercheur s'engage à remettre au promoteur ou à défaut au Directeur de l'Unité, les originaux de tout document, quelle qu'en soit la forme, relatifs aux expériences réalisées et aux résultats obtenus. Le Chercheur peut cependant en garder une copie.

## **7. Répartition des revenus de la valorisation**

### **7.1. Revenus de la valorisation**

Au sens du présent règlement, on entend par Revenus de la valorisation, tout type de revenu octroyé aux FUNDP en échange de la cession ou la concession des Résultats des recherches. Les types de revenus concernés par cette disposition sont notamment les redevances liées aux ventes

ou aux résultats d'exploitation, les montants liquidés de manière forfaitaire, les dividendes et les plus-values réalisées sur les participations dans une société, cessionnaire ou concessionnaire.

**7.2.** Les Revenus de la valorisation revenant aux FUNDP sont affectés en premier lieu au remboursement des sommes directement engagées par les FUNDP pour la protection et la valorisation des Résultats des recherches, augmentées d'une marge de 25% au titre de contribution aux frais généraux liés à la protection et à la valorisation des Résultats des recherches. Sur base annuelle, la Cellule de Valorisation de l'Administration de la Recherche des FUNDP informera le Chercheur des sommes engagées par les FUNDP pour la protection et la valorisation et des revenus relatifs à la valorisation des Résultats des recherches.

**7.3.** Une fois ce remboursement effectué, le solde des revenus de la valorisation est réparti entre :  
- le patrimoine non affecté des FUNDP ;  
- le laboratoire (ou l'unité de recherche) dans lequel les Résultats des recherches ont été obtenus ;  
- à titre personnel et dans le respect des dispositions fiscales en vigueur, le (ou les) Chercheur(s) ayant obtenu les Résultats des recherches et mentionné(s) dans le formulaire de déclaration visé à l'article 5.1.  
Par défaut, la répartition s'effectue en parts égales.  
Toute autre répartition doit recueillir l'accord des parties concernées.

## **8. Dispositions finales**

Toute question dans l'interprétation ou l'application du présent règlement sera soumise à la médiation de l'Administrateur en charge de la recherche.  
En présence d'un conflit persistant n'ayant pu être résolu à l'amiable à l'aide de la médiation de l'Administrateur en charge de la recherche, les parties soumettront leur conflit à un arbitre indépendant désigné de commun accord. La décision de l'arbitre liera les parties de manière définitive.